The Dissertation Committee for Matthew Richard Sorenson Certifies that this is the approved version of the following dissertation:

Exploring the eukaryotic gene expression machinery using a single-cell yeast gene expression reporter

Committee:

Scott Stevens, Supervisor

Andrew Ellington

Vishwanath Iyer

Arlen Johnson

Edward Marcotte

Exploring the eukaryotic gene expression machinery using a single-cell yeast gene expression reporter

by

Matthew Richard Sorenson, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2014

Dedication

To my wife Hayley, who has loved, sacrificed and encouraged much throughout this journey.

To my son Max, who continually teaches me how to find joy in the “little things.”

To my parents, for the endless support and love you have shown me.

Acknowledgements

This body of work would have not been possible without the support and direction of my supervisor Scott Stevens. The past six years in his lab have been complete with times of success and times of failure. I appreciate his support throughout it all. I would also like to thank Arlen Johnson for not only allowing me to rotate in his lab, but for thoughtful questions and advice throughout my graduate career. Additionally I thank Drs. Vishy Iyer, Edward Marcotte and Andy Ellington for helpful advice and guidance as members of my dissertation committee.

I would like to thank the members of the Stevens and Johnson laboratories who have encouraged and supported me during my time. This includes, but is not limited to Sujin Lee, Jennifer Hennigan, Rea Lardelli, Richa Sardana, and Champ Gupton. I thank Jon Laurent for his help as a rotation student in the initial stages of the *SWM2* screen. My thanks also go out to Ashwini Devkota and Eun Jeong Cho of Ti3D for guidance and help with the small molecule screen. Cheers to Drs. Joshua Kritzer and Jason Gavenonis at Tufts for providing the cyclic peptide libraries and their helpful advice. Richard Salinas of the ICMB core facilities was instrumental in helping me with flow cytometry, as was Cecil Harkey with fluid robotics. I would also like to thank Mariana Grenadier of the College of Natural Sciences for help in graphic design.

Exploring the eukaryotic gene expression machinery using a single-cell yeast gene expression reporter

Matthew Richard Sorenson, Ph.D.

The University of Texas at Austin, 2014

Supervisor: Scott Stevens

It has become increasingly evident that gene expression processes in eukaryotes involve communication and coordination between many complex, independent macromolecular machines. To query these processes and to explore the potential relationships between them in the budding yeast *Saccharomyces cerevisiae*, we designed a versatile reporter employing multicolor high-throughput flow cytometry. Due to its design, this single reporter exhibits a distinctive signature for many defects in gene expression including transcription, histone modification, pre-mRNA splicing, mRNA export, nonsense-mediated decay, and mRNA degradation. Analysis of the reporter in 4967 non-essential yeast genes not only revealed striking phenotypic overlaps between similar functions, but also provided us a dataset in which to compare additional genetic or chemical perturbations. Utilizing a binning and clustering algorithm we developed we are able to compare reporter expression data from thousands of

mutants in a semi-automated fashion, essentially grouping mutants or conditions based on the levels of reporter signal.

I further utilized our reporter to screen a vast library of *in vivo* produced cyclic peptides using fluorescence-activated cell sorting (FACS), identifying a cyclic peptide that resulted in mild gene-specific pre-mRNA splicing inhibition. Additionally, I adapted our reporter assay to perform a high-throughput small molecule screen to identify inhibitors or modulators of specific gene expression processes. Our efforts led to the identification of a small molecule that inhibits pre-mRNA splicing in a dose-dependent manner. Moreover, I utilized our reporter system to quickly identify loss-of-function mutants in the poorly characterized gene *SWM2*. The three mutants of interest have reduced interaction with Tgs1p, the conserved trimethylguanosine synthase, which we believe leads to decreased hypermethylation of the 5' caps of spliceosomal snRNAs. Altogether this work describes the development, validation and utility of the versatile gene expression reporter system we developed, providing our lab and others a valuable tool to interrogate a wide-range of gene expression processes in yeast.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1: Introduction

1.1 Eukaryotic gene expression

Within the information-rich genomes of all living organisms lie the instructions for each cell to function correctly, as well as the template for generating future generations. In eukaryotic organisms, the double-stranded DNA (deoxyribonucleic acid) that constitutes the genome is tightly packed by proteins within the nucleus, forming a structure called chromatin (Wolffe 1998). Although the length of DNA is condensed by orders of magnitude, the genetic material is spatially constrained and gene expression is temporally controlled, resulting in a highly functional nuclear architecture that allows for regulatory mechanisms (Cremer and Cremer 2001; Misteli 2005; Herrmann 2013).

The process by which genetic information is decoded and used by the cell to generate a functional product is often referred to as gene expression. The flow of genetic information, which is commonly referred to as the central dogma of molecular biology (Crick 1958, 1970), is typically thought to start with DNA, to RNA (ribonucleic acid), and end with protein by the way of transcription and translation respectively. In the years subsequent to Francis Crick's seminal work, it has been demonstrated that genetic information can flow in reverse and horizontal directions (reverse transcription and RNA replication respectively). Gene expression also involves the post-transcriptional and post-translational regulation of gene products with processes such as RNA interference and

ubiquitination among many others. Additionally, the functional product of gene expression is not limited to proteins, as many RNAs are functional, such as ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear RNA (snRNA). Since the work I present here was completed in the budding yeast *Saccharomyces cerevisiae,* most of the information provided will relate to this organism, with exceptions being noted.

1.2 TRANSCRIPTION AND MRNA PROCESSING

The first step in gene expression is the generation of the messenger RNA (mRNA) from template DNA in a processed called transcription, which is catalyzed by the enzyme RNA polymerase II (RNA Pol II) in the nucleus. As the pre-mRNA is being produced by RNA Pol II, additional machineries are recruited to the nascent mRNA mainly by virtue of the C-terminal domain (CTD) of RNA Pol II, which can act as a landing pad for mRNA processing factors, such as capping enzyme that co-transcriptionally attaches a 7-methylguanosine to the 5' end of the mRNA (Ghosh and Lima 2010). Not only does the cap function in stabilizing the mRNA, it also is important for downstream processing and gene expression events such as pre-mRNA splicing, mRNA export and translation. In addition to 5' capping and pre-mRNA splicing, almost all pre-mRNAs are cleaved and polyadenylated prior export from the nucleus. The cleavage of the nascent transcript and subsequent addition of a poly(A) tail are coordinated by cis-acting RNA elements (Chan et al. 2011; Tian and Graber 2012). In yeast, mutation of

3’-end processing factors result in nuclear poly(A) RNA accumulation and mRNA export mutations result in hyper-polyadenlyation of mRNAs (Brodsky and Silver 2000; Hilleren and Parker 2001; Hammell et al. 2002; Dower and Rosbash 2002; Dower et al. 2004). These findings highlight the importance and connection of both 3’-end processing and mRNA export for a proper gene expression program.

1.2.1 Pre-mRNA splicing

A common feature of eukaryotic pre-mRNAs is the presence of noncoding internal sequences called introns that are removed in a process called pre-mRNA splicing, which results in ligated coding sequences (exons). Intron removal is catalyzed by the spliceosome which is a multi-megadalton macromolecular machine comprised of five small nuclear ribonucleoprotein (snRNP) complexes, as well as many transacting factors (Will and Lührmann 2011). The five snRNPs include a small nuclear RNA (U1, U2, U4, U5 or U6) and its associated proteins. The aforementioned 5’ cap has long been known to stimulate pre-mRNA splicing (Krainer et al. 1984; Konarska et al. 1984; Edery and Sonenberg 1985; Fresco and Buratowski 1996; Schwer and Shuman 1996), highlighting the functional connections between mRNA processing events. In terms of conservation of the splicing machinery, proteomic analysis has revealed that 85% of yeast proteins are conserved in humans, providing strong evidence that the yeast spliceosome that we study in our laboratory embodies an important and evolutionarily conserved splicing machinery (Fabrizio et al. 2009).

1.2.1.1 Pre-mRNA splicing mechanism

The first reaction in the splicing mechanism involves the 2' hydroxyl of the conserved branchpoint adenosine nucleophilically attacking the phosphodiester bond of the 5' splice site. The second reaction involves the newly formed 3' hydroxyl of the 5' exon attacking the phosphate immediately upstream of the 3' exon at the 3' splice site. A simplified version of the step-wise splicing model is shown in Illustration 1.1. This model begins with the U1 snRNP base pairing with the 5' splice site (Legrain et al. 1988; Séraphin et al. 1988). At the branchpoint the U2 snRNP replaces Msl5p and Mud2p in an ATP-dependent step. After rearrangements within the U2 snRNP, the preassembled U4/U6•U5 tri-snRNP associates with the pre-mRNA substrate. By the way of protein and RNA rearrangements within and among the snRNPs and other transacting splicing factors, the spliceosome is then activated. Key characteristics of the activated spliceosome is the release of U1 by the ATP-dependent helicase Prp28p (Staley and Guthrie 1999), and the separation of U4 from U6 (Raghunathan and Guthrie 1998), permitting U6 to bind at the 5' splice site, replacing U1 (Lesser and Guthrie 1993). Additionally, the U5 snRNP binds to the 5' and 3' exons.

When the activated spliceosome is in the precise structural conformation, the RNA helicase Prp2p initiates the first transesterification reaction by gross rearrangement of the branchpoint-binding U2 snRNP (REA'S PAPER), yielding the free 5' (upstream) exon and the intron lariat intermediate. Through ATP-independent and ATP-dependent rearrangements, the 3' splice site and 3'

hydroxyl of the 5' exon are positioned in the active site (Schwer and Guthrie 1991; James et al. 2002; Schwer and Gross 1998). The second transesterification then can occur, yielding the products of pre-mRNA splicing, the spliced exons and the intron lariat. As with higher eukaryotes, the majority of yeast introns are removed cotranscriptionally and pre-mRNA splicing is sensitive to chromatin marks and transcriptional rates (Oesterreich et al. 2011), once again highlighting the interconnectivity of gene expression processes.

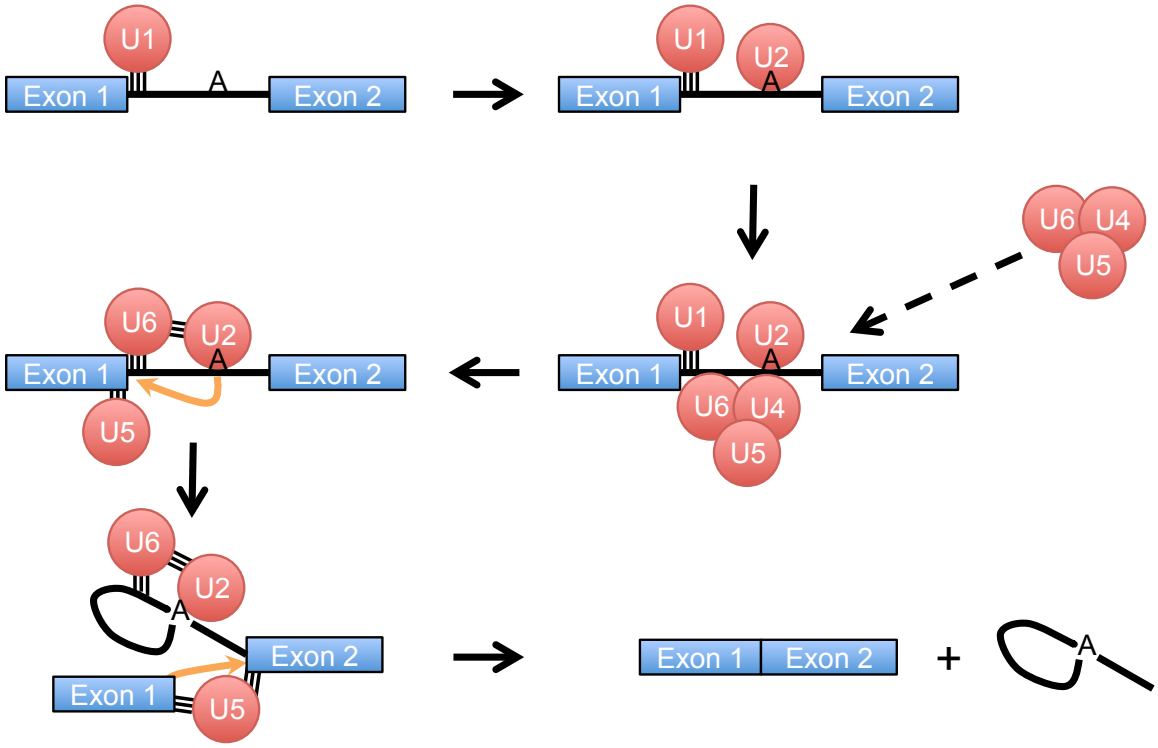


Illustration 1.1: Canonical pre-mRNA splicing model

The canonical model for a prototypical two exon (blue boxed) and one intron (black line in-between with the “A”) pre-mRNA. The pre-mRNA is first addressed by the U1 snRNP at the 5’ splice site followed by U2 at the branchpoint. After tri-snRNP (U4, U5 and U6) joins and rearrangements are made, U6 is bound at the 5’ splice site with U5 interacting with Exon 1. The first transesterification then occurs with the branchpoint adenosine (A) nucleophilically attacking the 5’s splice site (orange arrow). U5 bridges Exon 1 with the lariat intermediate and the second transesterification ensues with the 3’ hydroxyl of the Exon 1 attacking the last intron nucleotide, resulting in ligated exons and the intron lariat. Please see section 1.2.1 for more details.

1.2.1.3 Function and importance of introns

The yeast *Saccharomyces cerevisiae* contains relatively few and short introns, with only 5% of its 6000 or so genes containing a single intron (Spingola et al. 1999; Juneau et al. 2007; Parenteau et al. 2008). This being said, the majority of Pol II transcripts in yeast arise from intron containing genes, which mostly consist of highly expressed genes such as those encoding ribosomal proteins (Ares et al. 1999; Warner 1999). Additionally, the presence of introns most often results in increased levels of transcripts (Juneau et al. 2006). In contrast, 94% of human genes contain an intron, with an overall average of seven introns per gene (Lander et al. 2001; Venter et al. 2001; Sakharkar et al. 2004). In genes containing more than two exons, alternative splicing can occur, providing increased transcript and proteome diversity. As human introns range from 20 to 490,000 bp (Sakharkar et al. 2004), it may not be surprising that they contain other non-coding yet functional information and elements such as microRNAs, snoRNAs and Alu elements (Rearick et al. 2011; Hadjiargyrou and Delihas 2013).

1.3 mRNA EXPORT AND TRANSLATION

While mRNA biogenesis occurs within the nucleus, the resulting transcript must be exported to the cytoplasm to be translated. In higher eukaryotes the exon junction complex (EJC) and THO/TREX complex are deposited onto the mRNA through the processes of splicing (Bono and Gehring 2011; Luna et al.

2012). Moreover, polyadenlyation stimulates the nucleocytoplasmic transport of mRNAs (Huang and Carmichael 1996). In yeast, splicing is not believed to encourage mRNA export. The only known link between splicing and export in yeast is the SR protein Npl3p, which is a Mex67 adaptor, also enables cotranscriptional recruitment of U1 and U2 (Kress et al. 2008). Moreover Npl3p functions in histone H2B ubiquitination, providing a common link between mRNA transcription, splicing and export (Moehle et al. 2012).

Mex67p and the metazoan counterpart TAP/NXF1 are the transporters for nucleocytoplasmic transport and function to link poly(A) RNA to the nuclear pore FG-nucleoporins (Segref et al. 1997; Gruter et al. 1998; Kang and Cullen 1999; Herold et al. 2000; Zenklusen and Stutz 2001). Mex67p is loaded onto mRNAs cotranscriptionally and also facilitates anchoring of genes at nuclear pore complexes (Gwizdek et al. 2006; Dieppois et al. 2006). The main RNA adaptors for Mex67p/TAP are Yra1p in yeast or Aly/REF in higher eukaryotes (Strasser and Hurt 2000; Stutz et al. 2000; Zenklusen et al. 2001). Additionally the poly(A)-biding protein Nab2p (Iglesias et al. 2010), the THO/TREX complex which Yra1p associates with (Chávez et al. 2000; Libri et al. 2002; Strasser et al. 2002), the TREX-2 complex (Fischer et al. 2002; Rodriguez-Navarro et al. 2004; Wilmes et al. 2008; Jani et al. 2009; Ellisdon et al. 2010), and the previously mentioned Npl3p can all function in coupling mRNA biogenesis with export.

For mRNAs, as opposed to noncoding RNAs, translation is required for the functional product of the gene to be synthesized. As with the spliceosome,

the ribosome is a large dynamic machine comprised of both RNA and protein. The abovementioned 5' cap on mRNAs also functions to recruit translation machinery to the mRNA (Sonenberg 1993). Additionally, the 3' poly(A) tail binding protein Pab1p can functionally, but not covalently, circularize the mRNA by associating with the cap binding eIF4E and the translation initiation factor eIF4G, resulting in enhancement of translation (Gallie 1991; Tarun and Sachs 1995; Wells et al. 1998; Gross et al. 2003; Sonenberg and Hinnebusch 2009). Although the biosynthesis of proteins is of utmost importance to generate functional molecules, their function and stability can be regulated via posttranslational mechanisms, which is beyond the scope of this work.

1.4 mRNA DEGRADATION

As important as mRNA biogenesis and translation are in gene expression, the stability and degradation of mRNAs are crucial for controlling and regulating the overall result. When transcripts are being translated, they are relatively protected from the degradation machinery by virtue of the ribosome and the cap and poly(A) binding factors. Cytoplasmic mRNA degradation commonly occurs in structures known as P-bodies (Parker and Sheth 2007). The major degradation pathways most often begin with deadenylation of the 3' poly(A) tail by the Ccr4/Not complex (Tucker et al. 2002). Most often the decapping enzyme Dcp1p/Dcp2p then removes the 5' cap (LaGrandeur and Parker 1998; Dunckley and Parker 1999), followed by 5' to 3' degradation by Xrn1p (Larimer and

Stevens 1990; Larimer et al. 1992). It was recently demonstrated that defects in the cytoplasmic 5' to 3' decay machinery lead to transcription downregulation (Haimovich et al. 2013). Additionally, they show that these same components shuttle into the nucleus and associate with chromatin near transcription start sites and stimulate transcription initiation and elongation. This once again highlights the interconnectedness of gene expression process. Transcripts in the cytoplasm can also be degraded in a 3' to 5' direction by the exonuclease Dis3p within the exosome once the poly(A) tail is removed (Dziembowski et al. 2007).

Eukaryotic cells have also evolved machinery and mechanisms to purge the cell of potentially deleterious transcripts, such as one with a premature termination codon (PTC). The nonsense-mediated decay (NMD) pathway depends on detection of PTCs by the ribosome (González et al. 2001). NMD-targeted transcripts are most often decapped and shunted into the 5' to 3' decay pathway (Muhlrad et al. 1994; Hagan et al. 1995; Beelman et al. 1996), in contrast with the deadenylation/3' to 5' pathway of non-NMD targets. In addition to PTCs, extended 3' untranslated regions can also serve as a signal for NMD (Muhlrad and Parker 1999b, 1999a; Das et al. 2000). In addition to NMD, there are also no-go and non-stop decay pathways that handle transcripts with stalled translation elongation or without a termination codon respectively (Harigaya and Parker 2010; Vasudevan et al. 2002).

The last mRNA degradation pathway of note is the nuclear exosome and TRAMP (Trf4p/Air2p/Mtr4p Polyadenlyation) complex (Raijmakers et al. 2004;

LaCava et al. 2005; Houseley et al. 2006; Schmidt and Butler 2013). Within the nucleus the exonucleases Dis3p and Rrp6p function as a surveillance system that regulates levels of coding and noncoding RNAs, including intron-containing pre-mRNAs (Lemieux et al. 2011). It is interesting to note that the TRAMP complex and Rrp6p are cotranscriptionally recruited to intron-containing genes (Kong et al. 2014) and that Rrp6p escorts spliced and unspliced transcripts throughput their nuclear lifecycle, releasing them in the vicinity of the nuclear pore (Hessle et al. 2012). These connections between biogenesis and quality control once again highlight the interconnectedness of gene expression processes. There are of course additional means of mRNA degradation and regulation such as small interfering RNAs, microRNAs, localization and riboswitches; but these are beyond the scope of this work and will not be discussed.

1.5 DISSERTATION OBJECTIVES

Since joining the Stevens Laboratory, my efforts have been focused around developing and validating a novel reporter technology in order to study yeast gene expression in an efficient and meaningful way. The second chapter describes much of the design and validation of our reporter system, whereas the third and fourth chapters explore two different ways I utilized the reporter in my studies. The final chapter describes the overall significance of my work and the future directions that our reporter and key findings can be taken.

Chapter 2: Gene expression reporter design and validation

2.1 Background

Eukaryotic gene expression is orchestrated by a complex, dynamic, and regulated network of protein-protein and protein-nucleic acid interactions. These interactions require several multi-protein/ribonucleoprotein complexes working in concert to achieve and maintain proper levels of transcribed mRNA and translated protein (Komili and Silver 2008), and to avoid a disease state. From the organization of chromatin to the degradation of mRNAs, the ability to study the individual processes within gene expression, as well as the connections between them, has been aided by technological advances including the use of reporter genes that are sensitive to defects in a specific gene expression process. Many of these techniques and reporters have been developed for the model eukaryotic organism *Saccharomyces cerevisiae*.

2.1.1 Existing pre-mRNA splicing reporters

To query the pre-mRNA splicing pathway in yeast, initial efforts involved the fusion of intron-containing fragments to various reporter genes such as *lacZ* (Larson et al. 1983; Teem and Rosbash 1983; Jacquier et al. 1985; Rain and Legrain 1997), *HIS4* (Parker and Guthrie 1985), and a thymidine kinase (Fouser and Friesen 1986). These relied on the spliceosome to excise the intron from reporter gene pre-mRNA to produce a translation-competent mRNA, ultimately generating the measured signal. These reporters often had limitations of low

signal and a narrow dynamic range. A complementary reporter with splicing-dependent signals was also created in which a synthetic intron was cloned into the *lacZ* gene such that *β*-galactosidase activity resulted from translation of the unspliced pre-mRNA (Legrain and Rosbash 1989), resulting in increased signal under suboptimal pre-mRNA splicing conditions. This was unique because the construct reports levels of pre-mRNA splicing substrate and not product. Although this reporter improved upon some limitations, and has additional utility as an mRNA export assay, it still required extract preparation and biochemical analysis originating from one or more heterogeneous populations of cells.

Further progress was made with the development of the *ACT1-CUP1* reporter which was designed so that the degree of copper resistance positively correlated with pre-mRNA splicing efficiency (Lesser and Guthrie 1993). Although this reporter has proved to be instrumental in examining cis-pre-mRNA splicing elements and transacting factors, it does have limitations. Most notably the endogenous nonessential *CUP1* gene is often present in two or more copies in most laboratory strains (Karin et al. 1984; Tohoyama et al. 1996), requiring that all genomic copies of this gene be removed in order to specifically query pre-mRNA splicing efficiency. Additionally, after the high copy reporter is introduced into cells, the data derive from growth on solid media containing an ensemble population of cells. More recently a series of *lacZ* splicing reporters with competing splice sites was developed in yeast which enables the identification of factors that have a propensity to regulate alternative splicing (Dreumont and

Séraphin 2013). There are also many alternative splicing reporters for use in higher eukaryotes that utilize a combination of reporter genes, but as alternative splicing is far less prevalent in budding yeast, these are beyond the scope of this work.

2.1.2 Existing reporters for other gene expression processes

Reporters that query other gene expression events in yeast, such as transcription have also been developed. One such system employs the synthetic genetic array methodology (Tong et al. 2001, 2004) as well as a dual fluorescent reporter system as a genome-wide approach to discover regulators of a promoter of interest (Fillingham et al. 2009; Kainth et al. 2009; Kainth and Andrews 2010). This system greatly improved upon early hybrid yeast promoter reporters that utilized *lacZ* (Guarente and Ptashne 1981; Guarente et al. 1982) or genes for nutritional requirements (Titz et al. 2006). It is important to note that the product of some reporter genes can be very stable, confounding gene induction studies and requiring approaches such as exploiting a destabilized GFP as a transcription reporter (Li et al. 1998). Another challenge with reporters, especially in high-throughput studies when using β-galactosidase activity, is the need for cell lysis. This can be alleviated using reporter genes/signals that do not require lysis, or other approaches such as a secreted luciferase (Tochigi et al. 2010).

In surveying gene expression, one also needs to consider the role of pre-mRNA and mRNA degradation, not just their biogenesis, processing, and export.

For example, two groups developed plasmid-based nonsense-mediated decay (NMD) reporters that use *lacZ* gene fusions with premature termination codons (Leeds et al. 1991; Wang et al. 2001). In these systems, there is an increase in *β*-galactosidase activity in the absence of functional NMD. A similar approach was used with a lacZ construct that was used to assay the effects of cis-elements on transcript stability (Muhlrad et al. 1995). There are also NMD reporters that use GFP (Paillusson et al. 2005) and chemiluminescence (Boelz et al. 2006) in mammalian cells, one for *Giardia lamblia* that uses chemiluminescence (Chen et al. 2008), and one utilizing GFP and/or DsRed in *Drosophila melanogaster* (Metzstein and Krasnow 2006).

2.1.3 Significance and importance of single-cell analysis and flow cytometry

Although a clonal population of cells may be genetically identical, there are often high-levels of cell-to-cell variation across several parameters. Analysis or observations based on the population as a whole can mask differences in subpopulations or a small minority of cells, which can often be biologically significant (Johnston 2012). One main source of cell heterogeneity or variation is the stochastic or noisy nature of gene expression. The sources and importance of this gene expression variability have been recently reviewed (Sanchez and Golding 2013; Frank 2013). Single-cell analyses, such as flow cytometry are

used to tease apart and identify these potentially interesting phenotypic differences.

Although flow cytometry has been the workhorse for single-cell analyses for decades, the ever-increasing repertoire of fluorophores, coupled with advances in optics, fluidics, and software have resulted in large high-content flow cytometry datasets. For example, the HyperCyt Autosampler (IntelliCyt) boasts an analysis rate of under 2 seconds per sample (Edwards et al. 2001, 2007; Black et al. 2011), facilitating the generation of ever-larger datasets. Although there are some built-in analysis tools within high-throughput flow cytometry (HTFC) platforms, they commonly analyze bulk statistics of the entire population of events and do not take into account the cell-to-cell variability within and among multiple parameters, limiting the power of HTFC.

2.1.4 High-throughput flow cytometry analysis

Available analysis platforms did not meet our needs of comparing 2D flow cytometry plots which resulted from our specific reporter assay (Sorenson and Stevens 2014). For instance SiZer can identify statistically significant features within single parameter histograms then compares them using hierarchical clustering (Pratt et al. 2009). However, this analysis tool lacks the ability to compare more than one parameter at a time. Another tool automatically identifies subpopulations of interest within multiparameter data (Kalina et al. 2009), but does not allow for comparison between heterogeneous samples. With a similar

constraint, FLAME (FLow Analysis with automated Multivariate Estimation), as well as Hyperlog and FCOM (Petrausch et al. 2006), define and characterize discrete populations in high-dimensional flow cytometry data (Pyne et al. 2009). One of the most advanced algorithms for comparing multiparameter/high-dimensional data is SPADE (Spanning-tree Progression Analysis of Density-normalized Events) (Qiu et al. 2011). SPADE specializes in analyzing datasets that represent lineage progression and/or include rare cell types while comparing numerous functional markers simultaneously.

Although these analysis tools work well for the specific tasks for which they were designed, we required a tool that compared the shapes of 2D flow cytometry plots across a large number of samples allowing for variable parameter input.

2.2 MATERIALS AND METHODS

2.2.1 Construction of yeast strains

Standard yeast methods were used (Sherman 1991). Isogenic wild-type yeast strains BY4741 or BY4742 were used as wild-type comparisons except where noted. For the mCherry-Ufd2p-GFP strain (SS4065), a PCR product containing the *TDH3* promoter and the mCherry ORF was used to genomically N-terminally tag *UFD2-GFP* (SS4066) by homologous recombination (Huh et al. 2003) using the standard lithium acetate transformation method (Gietz and Woods 2002). For directed deletion mutant screen and the deletion collection

screen, the Yeast MATalpha Collection was used (Thermo/Open Biosystems; Cat# YSC1054). For *mex67-5* experiments the *mex67-5* strain SS4142 (Segref et al. 1997) was transformed with empty (SS4145) or *MEX67* plasmids (SS4146) and a *CUP1*-Reporter, selecting only for the introduced plasmids not the "native" *mex67-5* plasmid. Refer to Table 2.1 for a complete table of yeast strains used in this chapter, as well as Table 2.2 and Table 2.3 for plasmids and DNA oligonucleotides, respectively, used to make these strains.

Table 2.1 Yeast strains used in Chapter 2

Strain (SS)	Mating type	Genotype
4004	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* (BY4741)
4011	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* + pRS426 TDH3-mCherry (MS122) (mCherry only control)
4022	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* +pRS316 TDH3-GFP (MS125) (GFP only control)
4030	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* + pRS316 *TDH3*-BPS mutation Reporter (MS#135)
4031	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* + pRS316 *TDH3*-3'SS mutation Reporter (MS#136)
4034	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* + pRS416 (MS103) (Nonfluorescent control)
4050	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* (BY4742)
4052	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* + pRS316 *TDH3*-Reporter
4065	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 UFD2-GFP-His3MX*
4066	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 URA3::GPD-mCherry-UFD2-GFP-His3MX* (GFP-mCherry dimer control)
4142	a	*ade2,his3,leu2,trp1,ura3,mex67:HIS3* + pUN100 *LEU2 mex67-5* (AJY1231/PSY1687)
4145	a	*ade2,his3,leu2,trp1,ura3,mex67:HIS3* + pUN100 *LEU2 mex67-5* + pRS426 (MS107)+ pRS414 *CUP1*-Reporter (MS388)

(Table 2.1 continued)

4146	a	*ade2,his3,leu2,trp1,ura3,mex67:HIS3* + pUN100 *LEU2 mex67-5* + pRS426 *MEX67* (MS382)+ pRS414 *CUP1*-Reporter (MS388)

Table 2.2 Plasmids used in Chapter 2

MS #	Common Name	Purpose	Source
101	pRS414	Yeast Expression Vector	(Sikorski and Hieter 1989)
103	pRS416	Yeast Expression Vector	(Sikorski and Hieter 1989)
107	pRS426	Yeast Expression Vector	(Sikorski and Hieter 1989)
120	pRS316 *RPL25*-GFP	Yeast expression vector with eGFP	(Gadal et al. 2001)
122	pRS426 *TDH3*-mCherry	mCherry over expression control. Also used as template for dimer control	This work
124	Synthesized reporter fragment	Exon 1 through exon 2	Mr. Gene.
125	pRS316 *TDH3*-GFP	eGFP over expression control. Also used to clone reporter	This work. Plasmids MS 120 and 130
127	pRS316 *TDH3*-Reporter	*TDH3* driven reporter	This work. Plasmids MS 124 and 125
130	pGEM *TDH3*	Cloning vector with *TDH3* promoter	This work
135	pRS316 *TDH3*-BPS mut. Reporter	BPS mutant reporter	This work. Plasmid MS 127 and oligos MS 109-110
136	pRS316 *TDH3*-3'SS mut. Reporter	3'SS mutant reporter	This work. Plasmid MS 127 and oligos MS 111-112

(Table 2.2 continued)

162	TA *CUP1* Promoter	Cloning vector with *CUP1* promoter	This work.
163	pRS316-*CUP1*-Reporter	CEN *URA3* inducible reporter	This work. Plasmids 162 and 127.
251	pRS313 Reporter qPCR1-2-3	qPCR Control Template	This work.
283	p*GAL1-RRP6*	*RRP6* overexpression clone from Yeast ORF Collection	Thermo Scientific
382	pRS426 *MEX67*	Wild-type *MEX67* expression (pAJ282)	(Hung et al. 2008)
388	pRS414 *CUP1*-Reporter	CEN *HIS3* inducible Reporter	This work. Plasmids 101 and 163.

Table 2.3 DNA oligonucleotides used in Chapter 2

MS#	**Sequence (5' to 3')**	**Purpose**
109	CGATAAAATGTCCAGTTGAAAACCTGTTTTACTAGCGATTTAAAAATTGTATTTCAT	Site-directed PCR Mutagenesis – Reporter BPS A to G mutation FWD
110	ATGAAATACAATTTTTAAATCGCTAGTAAAACAGGTTTTCAACTGGACATTTTATCG	Site-directed PCR Mutagenesis – Reporter BPS A to G mutation REV
111	TTTCATTACAATATTTTTTTTGTACTGCCGGTTAAGGTCGTATCTCT	Site-directed PCR Mutagenesis – Reporter 3'SS A to T mutation FWD
112	AGAGATACGACCTTAACCGGCAGTACAAAAAAAATATTGTAATGAAA	Site-directed PCR Mutagenesis – Reporter 3'SS A to T mutation REV
146	GTGGGCGGCCGCTCCCATTACCGACATTTGGGCGCTA	PCR – *CUP1* Promoter FWD
147	GTGGGGATCCTGATTGATTGATTGATTGTACAGTTTG	PCR – *CUP1* Promoter REV
257	TCACGTCTCAGGCCGGTTAAGGTC	qPCR – Reporter mRNA FWD

(Table 2.3 continued)

258	GCAGCTTGCCGGTGGTGCAGATG	qPCR – Reporter mRNA REV
303	TAATAGAGGTGTACGCCGCAATCATTGTTAATCATTAAGTAATTGCCATTAGAGTGCACCATACCACAGC	PCR – mCherry-*UFD2*-GFP dimer FWD
304	GTATCGGAAGGATCGGTGGTGATTTGTAAAATATCTTCTATGGCGGTCATCTTGTACAGCTCGTCCATGC	PCR – mCherry-*UFD2*-GFP dimer REV
309	CGTTAAGCACCCAGCTGACA	qPCR – Reporter pre-mRNA FWD
310	TGCATAACTGGACCGTCAGATG	qPCR – Reporter pre-mRNA REV

2.2.2 Site-directed PCR mutagenesis

Primers were designed to both incorporate the desired mutation and to amplify the template plasmid. Reaction conditions were as described previously (Barnes 1994). PCR cycling was performed at 95°C 1 minute followed by 17 cycles of [95°C 50 seconds, 45°C - 65°C 50 seconds, 70°C (1.25 min/Kb of template)] followed by incubation at 70°C for one elongation-length step. Reactions were treated with the restriction enzyme DpnI to destroy the template plasmid and were transformed into competent *E. coli*. Mutant plasmids were screened by diagnostic restriction enzyme digest and sequencing.

2.2.3 Robotic high-throughput 96-well yeast transformation

We adapted the standard lithium acetate (Gietz and Woods 2002) and microplate methods (Gietz and Schiestl 2007) to utilize robotic techniques and increase transformation efficiency and speed. Except where noted, liquid

handling was performed using a BioMek FX laboratory automation workstation (Beckman Coulter). 96-well plates containing cryogenically preserved yeast strains were thawed and small volumes transferred to a 96-well plate containing 150 µL/well fresh YPD medium supplemented with 200 µg/ml G418 (effective concentration; Gibco) using sterile pin replicators. Strains were grown for 16-20 hours at 31°C and breathable sealing film (Axygen) was used to minimize cross contamination. Cells were pelleted and washed twice with 100 mM lithium acetate, then mixed with a pin replicator in pre-warmed transformation master mix (33.3% PEG 8000, 100 mM lithium acetate, 2 mg/mL salmon sperm DNA and 500 ng plasmid/reaction at 42°C). Plates were incubated at 42°C for 45 minutes, washed twice with ddH_2O and resuspended in 50 µL ddH_2O. To select for transformants, 20% of the cell suspension was spotted onto solid minimal selecting for the plasmid (SD-uracil + 200 µg/mL G418). The remaining 80% of cell suspension was cryogenically preserved by adding an equal volume of 50% glycerol and stored at -80°C. If transformants were not obtained in the initial 20% spot plating, the remaining transformation mixture was thawed and spread onto selective media.

2.2.4 Flow cytometry analysis

For standard flow cytometry, log-phase yeast were fixed in 10% formaldehyde, washed and resuspended in 1X PBS. For high-throughput flow cytometry, 96-well plates containing stocks of the deletion strains harboring the

reporter were replicated into 150 µL selective media. When the average well in a plate reached mid-log phase, cells were harvested, washed and resuspended in 100 µL 1X PBS and analyzed immediately with a Becton Dickinson LSRFortessa equipped with a high-throughput sampler unit. For excitation of eGFP and mCherry, 488 nm and 561 nm lasers were used respectively. For detection a 505 LP and 530/30 BP were used for eGFP and a 600 LP and 610/20 BP for mCherry. For each sample 21,000 events were collected and 10,000 events are shown in the phenographs. FlowJo (version 9.6.4, Tree Star Inc., Ashland OR) was used to analyze data and to prepare phenographs, plotting cells based on their fluorescent intensity (GFP on x-axis and mCherry on y-axis), which is in arbitrary units based on predetermined cytometer settings.

Data in Figure 2.12 were obtained from the dataset “T cell stimulation (human)-titration and timecourse” from the public Cytobank repository and submitted by A. Landrigan. These data represent precursors to the work presented in their subsequent publication (Landrigan et al. 2011). The same gating hierarchy (FSC-A/SSC-A (FSC/SSC), FSC-A/FSC-H, CD45RA/CD45RO) was used to identify live, single, naïve CD45RA+ cells respectively for further analysis. Briefly, human CD4+ T cells were purified and stimulated with anti-CD3 (OKT3 or UCHT; 0.3 or 0.8 µg/mL) with or without at 10 µg/mL anti-CD28 for 3, 6, or 10 minutes. Upregulation of pSLP76/S6 and pCD247 was assessed.

2.2.5 Binning and clustering analysis of flow cytometry data (CHUBACA)

FlowJo (version 9.6.4, Tree Star Inc., Ashland OR) was used to create all flow cytometry plots and to process flow cytometry data. For further analysis, scale values for two parameters were exported as text files, one file for each sample. The text files were processed using SAS software (Version 9.2, SAS Institute Inc.). First, the user sets the range of x- and y-axis values, the number of bins for each axis, as well as the total number of events to be analyzed. When the analysis begins, a log transformation is applied to the scale values and individual events are assigned a bin based on those two scale values and the other applied settings. Binning begins at the origin (lowest value for each axis) then proceeds up the y-axis. It then returns to the x-axis, shifting to the right one bin then proceeding up, continuing until the last bin in the top right of the plot (highest values for each axis). Sample/data files that do not meet the standards set by the user-defined settings are not incorporated into the final analysis and this exception is reported. After binning all of the events for each sample, the results are exported into a text file that reports the number of events in each bin (columns) for each sample (rows). Data are then hierarchically clustered with Cluster (Eisen et al. 1998) and a dendrogram is prepared using Java TreeView (Saldanha 2004). I refer to our analysis pipeline as “CHUBACA” (Comparative Hierarchy Using Binning And Clustering Analysis).

When using CHUBACA I recommend naming samples in a simple and systematic format and then renaming the samples after the binning analysis

using a VLOOKUP function in Microsoft Excel. Per SAS limitations, sample/file names for TXT files should not contain any spaces or punctuation, except underscores. The character limit for samples is 12. It is recommended that TXT files be processed in smaller groups if working with a large dataset. For example, we broke up our analysis of the 5000 yeast deletion mutants into groups of eight 96-well plates (768 samples). The SAS source code is available upon request.

2.2.6 Total yeast RNA isolation

Two OD units of log phase cells were harvested by centrifugation in a microfuge tube and total RNA was extracted with 266 µL acid phenol and 266 µL AES Buffer (50 mM sodium acetate pH 5.3, 10 mM EDTA and 1% SDS). Samples were vortexed well and incubated for 6 minutes at 65°C, vortexing every minute (Wise 1991). Samples were placed in ice water for 5 minutes, then transferred to a Phase Lock Gel Heavy 2 mL tube (5 PRIME). Tubes were centrifuged for 5 minutes at 2500 rpm. Acid phenol (266 µL) was added; tubes were gently inverted five times and then centrifuged. Chloroform (266 µL) was added; tubes were inverted to mix and then centrifuged. The aqueous phase above the gel matrix was transferred to a new microfuge tube. 3 M sodium acetate (1:10 volumes, pH 5.3) was added, followed isopropanol (1.1 volumes). Tubes were incubated at -20°C overnight to precipitate the RNA. RNA was harvested and washed twice with 1 mL 70% cold ethanol and resuspended in 30-50 µL DEPC-treated water.

2.2.7 RT-qPCR analysis

Total RNA (5 µg) was DNase treated (5 U RQ1 DNase, Promega) for 20 minutes at 37°C. RNA Clean and Concentrator kit (Zymo) was used to purify the RNA per the manufacturer's protocol (6.5 µL DEPC water was used to elute, resulting in an final volume of 5 µl). To the DNase-treated RNA, the following was added: 1 µL 10X RT Buffer (0.5 M Tris-HCl pH 8.5), 1 µL random nonamer primers (25 µM; Integrated DNA Technologies; MS299-302) and 2 µL DEPC H_2O. The RNA/primer mix was incubated at 60°C for 5 minutes and immediately transferred to ice for 5 minutes. 11 µL of an RT master mix (1X RT Buffer, 3 mM $MgCl_2$, 10 mM DTT, 0.5 mM dNTPs and 5 U/µL MultiScribe RT enzyme) was added. The RT reaction was incubated overnight at 42°C and the resulting cDNA was diluted with four volumes DEPC H_2O. Each qPCR reaction contained 5 µL cDNA, 7.5 µL Power SYBR Green PCR Master Mix (ABI) and 2.5 µL primers (1.5 µM each). Standard thermocycling, fluorescence detection and analysis (standard curve) were completed using the ViiA 7 Real-Time PCR System (Life Technologies).

2.3 RESULTS

2.3.1 Design, cloning, and validation of the eukaryotic gene expression reporter

We designed a reporter that is amenable to high-throughput single-cell analysis and simultaneously generates signals reflecting pre-mRNA and mRNA levels. To accomplish our goal, we chose to drive transcription of our reporter with the strong and constitutive *TDH3* (*GPD*) promoter. The exons (EX1 and EX2 in Figure 2.1A) and intron were based on the relatively inefficiently spliced yeast *RPL28/CYH2* gene (Swida et al. 1986) to increase the chances that the unspliced transcript levels were detectable.

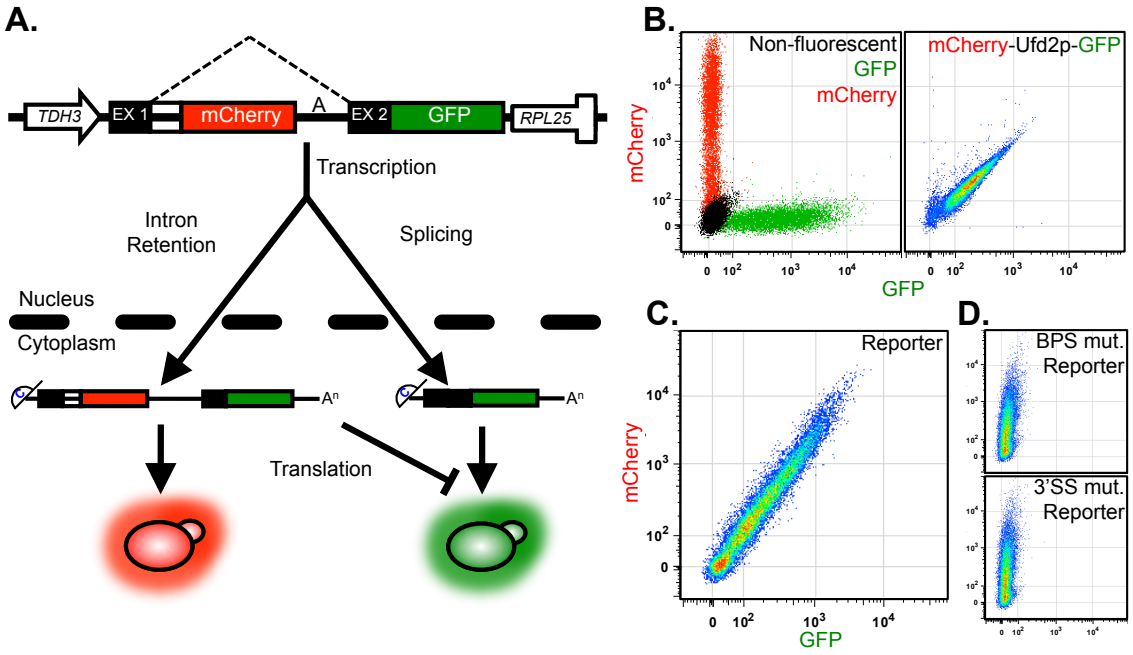


Figure 2.1: Gene expression reporter design and validation.

Design (A) of the gene expression reporter construct is shown. Expression of the synthetic reporter was driven by the strong constitutive *TDH3* promoter. Exon 1 (EX1), the intron and exon 2 (EX2) were designed using features of the inefficiently spliced *CYH2* gene. Dashed lines indicate the region comprising the intron. The mCherry ORF was placed in frame with EX1 and the eGFP ORF was cloned in frame with EX2. Exported intron-containing pre-mRNAs and mature mRNAs are translated resulting in proteins which fluoresce red and green respectively. The construct was cloned into pRS316 (see Materials and Methods for more detail). Control samples (B) for flow cytometry are shown. Phenographs for wild-type yeast harboring an empty vector (non-fluorescent), or vectors expressing GFP-only or mCherry-only are shown (left panel). In the right panel, a phenograph of the expression of the mCherry-Ufd2p-GFP dimer results in a 1:1 mCherry:GFP signal. The pseudo color represents cell density within the phenograph. Reporter expression in wild-type yeast (C) is displayed and referred to as a phenograph. Phenographs for branch point (A to G) and 3' splice site (CAG to CuG) mutant reporters (D) in wild-type yeast are shown (top and bottom panels respectively). The y- and x-axis depict mCherry and GFP fluorescence respectively.

An annotated version of the reporter sequence can be found in Appendix I. All nucleotide positions and descriptors below refer to Appendix I. To generate the reporter, the region from EX1 through EX2 (640-1598) was commercially synthesized (Mr. Gene; Regensburg Germany) and cloned into the BamHI site of pRS316-*TDH3*-eGFP (derived from pRS316-*RPL25*-GFP) by replacing the *RPL25* promoter and ORF with the *TDH3* promoter (31-639) using BamHI and NotI restriction enzymes. In the design, we removed an inhibitory sequence (Swida et al. 1988) within the *CYH2* intron and positioned a yeast codon optimized mCherry ORF (nucleotides 763-1470 in the sequence below) upstream of the branch point sequence (1526-1532) which removed 32 bp from the intron. The mCherry ORF was placed in frame with EX1 and a stop codon was placed immediately at the end of the mCherry ORF. The resulting plasmid contains a small portion of the *CYH2* EX2 fused in frame with the eGFP ORF (1600-2322). This construct was cloned into a suite of pRS vectors (Sikorski and Hieter 1989) using BamHI/SalI, maximizing utility. In Appendix I, underlined bases are those that were mutated for the branch point mutant and 3' splice site mutant and the 5' GT and 3' AG intron boundaries are in lower-case text.

In this system, levels of exported pre-mRNA (intron containing) are reflected by red fluorescence via a codon-optimized open reading frame (ORF) for mCherry (Shaner et al. 2004) that was placed between the 5' splice site and branch point sequence, in frame with the first exon and a portion of the intron (Figure 2.1A left side). After nuclear export, translation of this pre-mRNA

terminates at a stop codon located at the end of the mCherry ORF. To measure the level of exported spliced message, the enhanced GFP (eGFP) ORF (Heim et al. 1995) was cloned in frame with the second exon (Figure 2.1A right side). When the pre-mRNA is unspliced, the second exon was designed to be out of frame with the first exon and a stop codon is present after the mCherry ORF, ensuring that GFP translation and fluorescence results only from the spliced message. The final construct was built into a series of low-copy centromeric plasmids with different auxotrophic and dominant selection markers to maximize flexibility.

To rapidly measure the levels of reporter fluorescence in individual yeast cells we utilize flow cytometry. In Figure 2.1B, we show flow cytometry data for an empty vector control (non-fluorescent) and cells over-expressing either GFP or mCherry alone (Figure 2.1B, left panel). As an additional control to determine the behavior of mCherry and GFP when present at high levels and at a defined 1:1 ratio, we designed a strain in which the yeast gene *UFD2* was tagged at the C-terminus with GFP (Huh et al. 2003) and was modified to contain the *TDH3* promoter followed by an N-terminal mCherry tag. This non-essential yeast gene was chosen as a linker between the fluorescence proteins as the X-ray crystal structure of this protein shows that the N- and C-termini are ≈146 Å apart (Tu et al. 2007), eliminating potential fluorescence resonance energy transfer between GFP and mCherry (Albertazzi et al. 2009), while keeping their mRNA and protein levels equal. Expression of the dimer resulted in equivalent amounts of

fluorescence in the flow cytometer, demonstrating that fluorescence operates as a suitable proxy for in vivo reporter transcript levels (Figure 2.1B, right panel).

In wild-type yeast cells, our reporter exhibited 2.5-fold greater red fluorescence than green fluorescence, indicating that while translating pools of pre-mRNA and mRNA are not 1:1, as in the dimer control, our reporter produces both translated pre-mRNA and mRNA at adequate levels for detection and with sufficient dynamic range available to measure changes (Figure 2.1C). We refer to the flow cytometry data presented in this work as a phenograph, as it graphically represents the reporter expression phenotype in a given yeast strain. To ensure that no GFP signal results from the pre-mRNA transcript, we mutated the intron branch point nucleotide from A to G or the 3' splice site from AG to UG. As predicted, both of these mutations abolished green fluorescence, demonstrating that intron removal is required for the production of GFP from this reporter (Figure 2.1D).

We note that the elongated shape of the phenograph is due to heterogeneity of cell volume as we observed a strong correlation between fluorescence intensity and cell volume (Figure 2.2A). Additionally, the fact that the cells fall near the middle of the phenographs allows for detection of an increase or decrease of at least two orders of magnitude with either signal. Reassuringly, the levels of fluorescent protein correlate with transcript levels by RT-qPCR (Figure 2.2B).

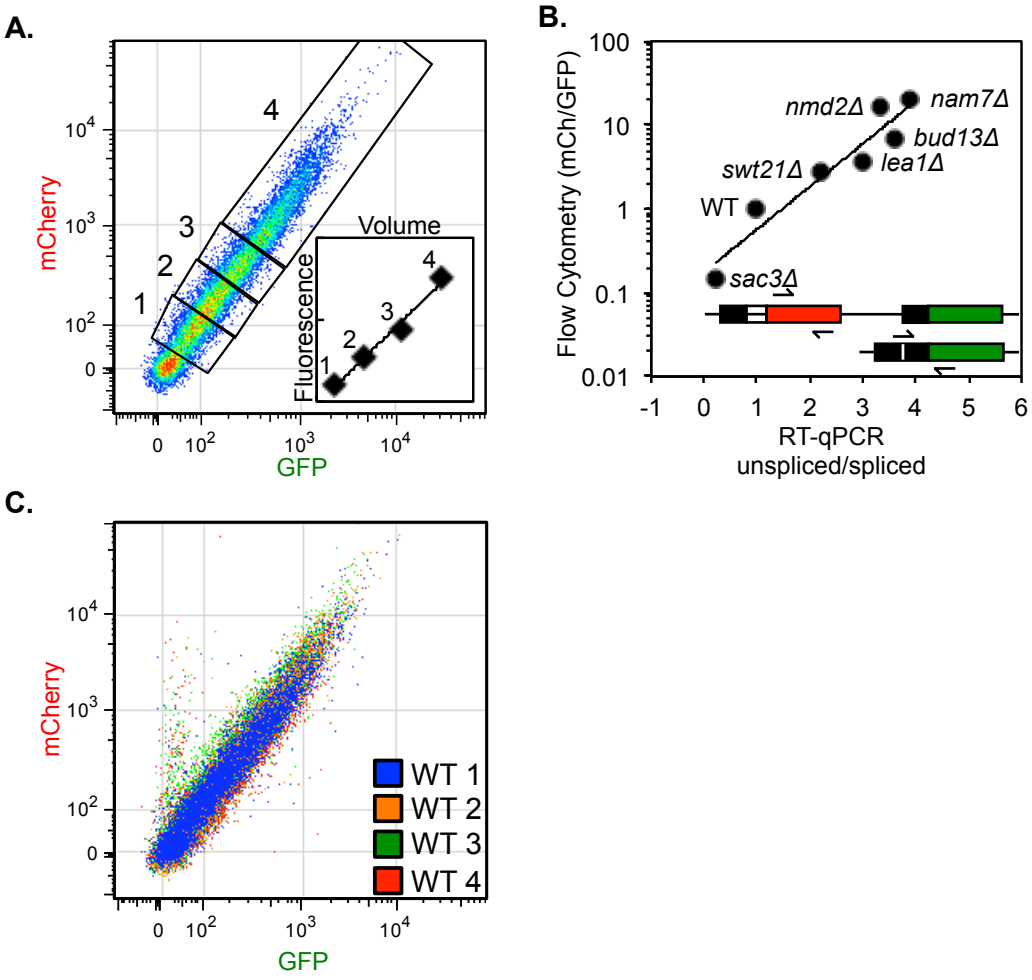


Figure 2.2: The reporter phenograph is a reproducible product of cell volume and reporter transcript levels.

Reporter expressing cells were gated (A) into four equally populated groups based on their position within the phenograph, moving out from origin (essentially grouping cells based on their fluorescence intensity). The inset compares the average total fluorescence with the average cell volume (forward scatter) for the four groups. An exponential regression line is applied (R^2 = 0.98). The flow cytometry mCherry/GFP ratio is plotted for several deletion strains versus the ratio of unspliced and spliced reporter transcripts (B), which was determined by RT-qPCR analysis. Flow cytometry and RT-qPCR ratios are relative to wild-type (WT) and an exponential trend line is shown (R^2 = 0.91). The inset cartoon depicts the primer location for the unspliced (top) and spliced transcripts. Phenographs from wild-type yeast analyzed on different days (C) are visually indistinguishable when overlaid.

As with any technology or assay, reproducibility is of critical importance, especially if one wishes to compare data from experiments performed on different days. To demonstrate this quality of our reporter, I compared data from wild-type cells containing our constitutive reporter from four separate experiments on different days. When these data are overlaid, they are virtually indistinguishable (Figure 2.2C), demonstrating great day-to-day reproducibility.

2.3.2 Deletion of genes within specific gene expression processes exhibit similar phenograph patterns

To assess the behavior of the reporter I generated phenographs from a small collection of yeast strains, each lacking a single gene functioning in some aspect of gene expression. It was expected that defects in the splicing machinery would decrease the efficiency of intron removal (Vijayraghavan et al. 1989), resulting in populations of cells that are red-shifted, due to a decrease in green fluorescent signal which would result from decreased levels of spliced reporter transcript. In the first set of splicing mutants (*mud2*Δ*, swm2*Δ*,* and *swt21*Δ) we detected a relatively minor red shift, whereas in the *bud13*Δ*, ist3*Δ*,* and *lea1*Δ strains we detect a more severe red-shift (cf. Figures 2.3A and 2.3B respectively). This shift in the ratio of the translation products of the pre-mRNA to mRNA is also reflected in the corresponding RNA levels as determined by RT-qPCR (Figure 2.2B).

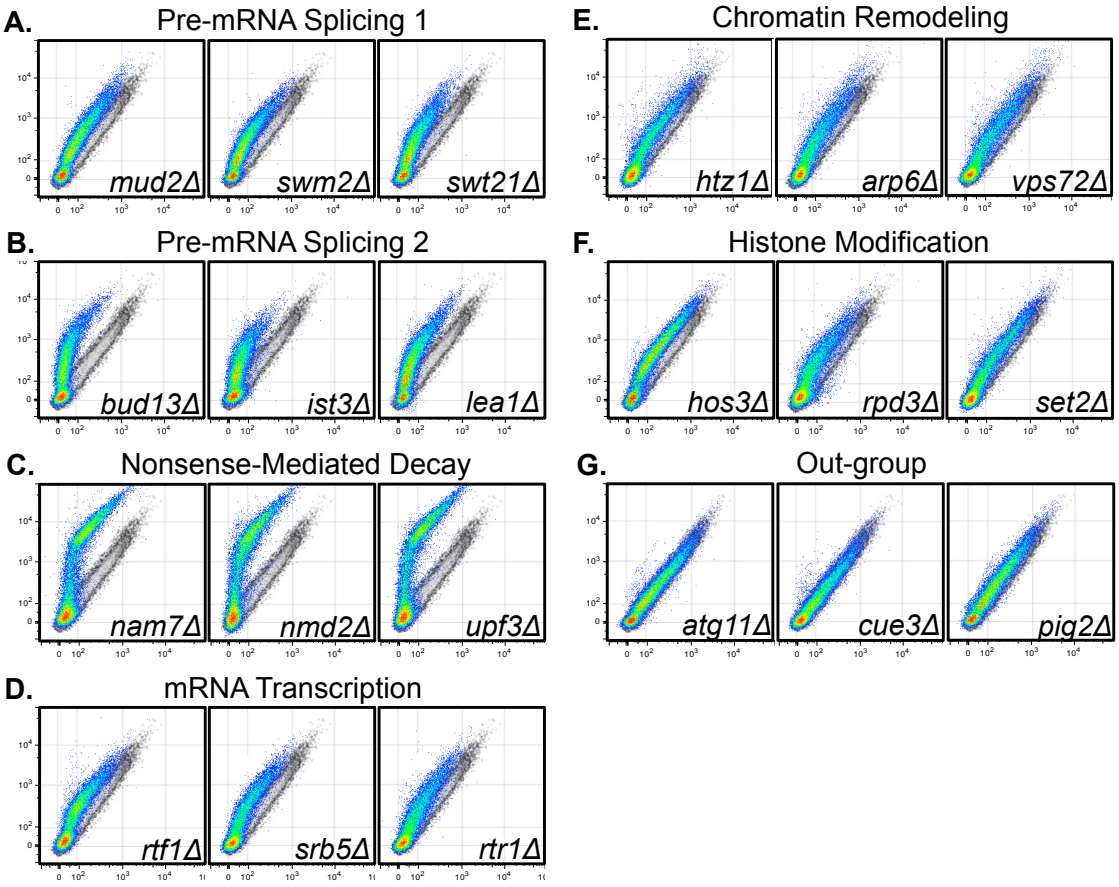


Figure 2.3: Mutations within specific gene expression pathways result in red-shifted phenographs.

The gene expression reporter described in Figure 2.1A and 2.1C was transformed into the indicated strains. Each phenograph overlay contains data from the indicated deletion strain in pseudo color and an isogenic wild-type strain (SS4050/BY4742) in the grayscale for comparison. Each gene expression process (A-F) is indicated above the phenographs and the out-group (G) represents three unrelated, randomly chosen deletion mutants to serve as negative controls. The y- and x-axis depict mCherry and GFP fluorescence respectively.

Next, we considered whether our reporter was sensitive to and behaved differently in strains containing mutations in other gene expression pathways. The most striking were the nonsense-mediated decay (NMD) mutants *nam7*Δ*,* *nmd2*Δ and *upf3*Δ (Figure 2.3C; He et al. 1997). This is not surprising because the long 3' UTR of the reporter pre-mRNA, which extends from the 3' end of the mCherry ORF through the GFP coding region, is predicted to be an ideal substrate for NMD in wild-type yeast (Muhlrad and Parker 1999). Without a functional NMD pathway, we would expect increased mCherry signal due to increased stability and translation of the reporter pre-mRNA. As predicted we detected a severe red-shift which was a result of over a 10-fold increase in red fluorescence with very little change in green fluorescence intensity, supporting the suggestion that disruptions in the NMD pathway result in accumulated levels of translatable reporter pre-mRNA transcripts (He et al. 1993). RT-qPCR analysis further corroborates this, revealing a large increase in the pre-mRNA/mRNA ratio in NMD mutants compared to wild-type cells (Figure 2.2B). NMD has also been shown to mask intron accumulation of poorly spliced transcripts by degrading the pre-mRNA transcripts in yeast (Sayani et al. 2008), which is also consistent with the red-shifted phenographs and RT-qPCR results. Our results indicate that the reporter pre-mRNA is inefficiently spliced and that this is masked by NMD in wild-type cells, similar to the suboptimally spliced natural introns highlighted by Sayani et al. I also analyzed forward scatter values in these mutants to see if cell volume increased compared to wild-type. There was no difference seen in these

mutants, supporting the conclusion that increased translation of the stabilized pre-mRNA resulted in the red-shift and not an accumulation of fluorescent proteins in a larger cell (data not shown).

Defects within early steps of gene expression also altered reporter expression. I analyzed the phenographs of several mutants lacking factors involved in mRNA transcription (*rtf1*Δ, *srb5*Δ and *rtr1*Δ; Figure 2.3D), chromatin remodeling (*htz1*Δ, *arp6*Δ and *vps72*Δ; Figure 2.3E) and histone modification (*hos3*Δ, *rpd3*Δ and *set2*Δ; Figure 2.3F). The phenographs from these mutants were primarily red-shifted, but to a lesser degree than in NMD mutants and more closely resembling some pre-mRNA splicing mutants (Figure 2.3A). The observed red-shift in these mutants, which is a result of a decrease in green fluorescence (spliced signal), is not surprising due to the large number of reports describing the interconnected nature of chromatin remodeling, histone modification, transcription and pre-mRNA splicing. We also note that the phenographs for mutants with no direct function in gene expression do not significantly change compared to wild-type (*atg11*Δ, *cue3*Δ and *pig2*Δ; Figure 2.3G).

Interestingly, a subset of gene expression factor deletions produced an overall green-shifted phenograph rather than red-shifted phenographs, indicating a relative increase in the spliced reporter signal and/or a decrease in pre-mRNA signal. This was most striking in yeast strains lacking factors of the TREX-2 complex (*sac3*Δ, *thp1*Δ, *sus1*Δ, and *sem1*Δ; Figure 2.4A), which couples mRNA

transcription and mRNA export (Fischer et al. 2002; Rodriguez-Navarro et al. 2004). The loss of TREX-2 core components Sac3p or Thp1p results in indistinguishable phenographs, which would be expected. Sus1p has been proposed to serve as a link between the SAGA complex and TREX-2 complex (Ellisdon et al. 2010), which results in a similar but not identical phenograph when removed from yeast. Analysis of different classes of mRNA export related mutants, such as the THO complex and components of the nuclear pore (Stewart 2010), demonstrate a similar green-shift (Figures 2.4B-C). Possible explanations for this effect are changes in mRNA stability, a preferential export of specific transcripts, or as our RT-qPCR data suggests, the defect in mRNA export likely leads to a lengthened nuclear residency and might increase the likelihood of splicing, increasing the green to red ratio (Figure 2.2B). It is also important to note that others have used translationally competent reporter pre-mRNAs as mRNA export reporters, providing further support for the green-shifted effect we observed (Galy et al. 2004; Dufu et al. 2010; Teplova et al. 2011; Caporilli et al. 2013).

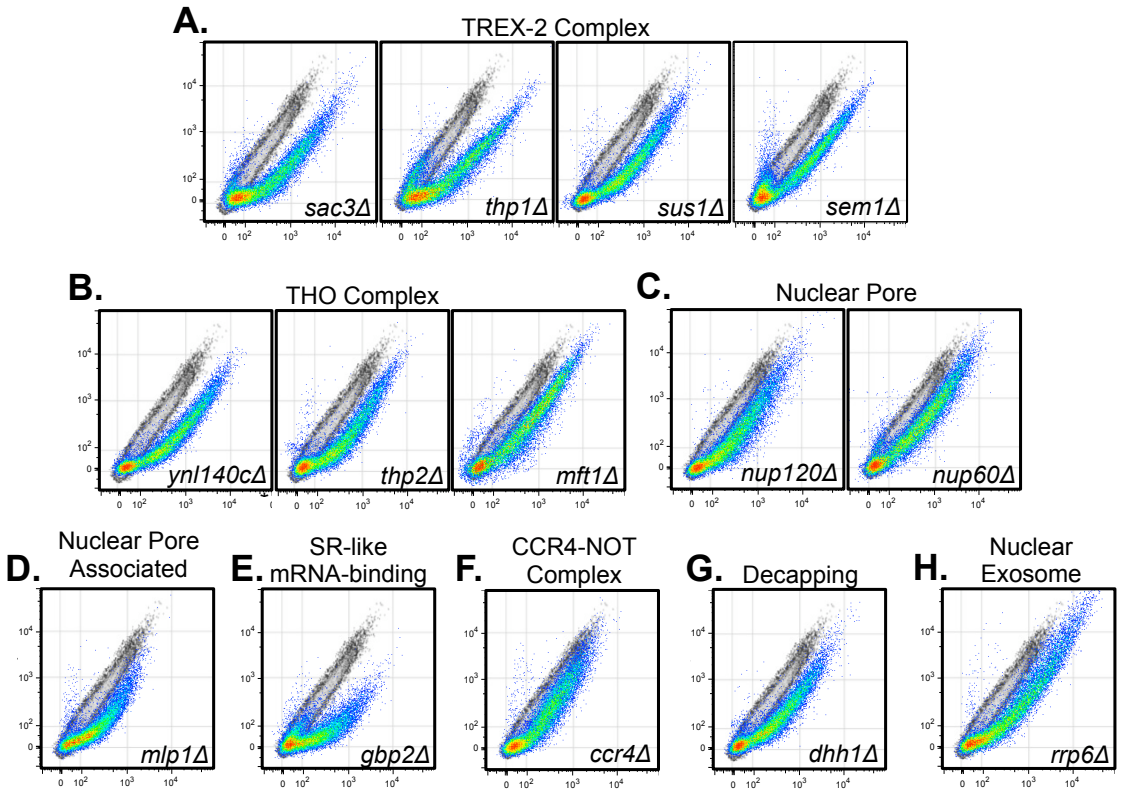


Figure 2.4: Mutations within specific gene expression pathways result in green-shifted phenographs.

The gene expression reporter described in Figure 2.1A was transformed into the indicated deletion strains. Each phenograph overlay contains data from the indicated deletion strain in pseudo color and an isogenic wild-type strain (SS4050/BY4742) in the grayscale for comparison. Each gene expression process or complex (A-H) is indicated above the phenograph. Asterisk indicates a deletion that overlaps the verified ORF *THO2*. The y- and x-axis depict mCherry and GFP fluorescence respectively.

One interesting green-shifted phenograph was that of the *mlp1*Δ mutant (Figure 2.4D), a result of a nearly 80% decrease in mCherry signal, with a minor change in GFP. This is in contrast to the green-shifts observed in the previously described mutants (Figures 2.4A-C, which primarily result from an increase in GFP). Mlp1p has been shown to mediate retention of unspliced reporter mRNAs, resulting in increased pre-mRNA reporter translation in its absence (Galy et al. 2004; Hackmann et al. 2014), which is in contrast with the result shown here. Although both reporters are weakly spliced due to suboptimal intron architecture, our reporter contains the extended 3'UTR that is efficiently degraded in the cytoplasm via NMD (Figure 3C). This feature, along with the role of Mlp1p as a checkpoint for export competent messenger ribonucleoparticles (mRNPs; Green et al. 2003; Palancade et al. 2005), leads us to believe that exported unspliced reporter mRNPs in *mlp1*Δ cells could be improperly packaged, leading to a decreased signal by virtue of NMD and/or other degradation pathways. Further studies are needed to confirm this. Of note, I also observed a similar effect in cells lacking the SR-like protein Gbp2p (Figure 2.4E), which was recently shown to be involved with Mlp1p in a potential mRNP quality control mechanism coupled with export (Hackmann et al. 2014).

We also note that the deletion of factors involved the turnover of mRNA transcripts, which stabilizes mRNA messages, also results in green-shifted cells (*dhh1*Δ and *ccr4*Δ; Figures 2.4F-G; Goldstrohm and Wickens 2008). Although the *rrp6*Δ mutant results in an overall green-shift like *dhh1*Δ and *ccr4*Δ (Figure 2.4H),

it is important to note that the *rrp6*Δ mutant is the only deletion mutant that results in significant increases of both GFP and mCherry fluorescence. Rrp6p is a nuclear exosome component with 3'-5' exonuclease activity, and is thought to degrade aberrant transcripts, among other functions (Chlebowski et al. 2013). Loss of this enzyme could stabilize the pre-mRNA and in turn increase the probability of it being spliced within the nucleus. This may lead to increased signals of both reporter messages. On the other hand, the *dhh1*Δ and *ccr4*Δ mutants only increase in the green fluorescence spliced signal, which is most likely due to their function in decreasing the stability of mature cytoplasmic transcripts in a wild-type strain.

I also looked into the effects on total fluorescence of many of these mutants by plotting the mean mCherry and GFP values for the 21,000 cells that were analyzed by flow cytometry (Figure 2.5). As expected, we observed a decrease in the overall fluorescence in mRNA transcription mutants, as well as specific patterns for other mutant groups, such as NMD and mRNA export. The directed screen of gene expression mutants demonstrates that our reporter is uniquely sensitive to many gene expression events, increasing its utility by permitting numerous biological pathways to be queried at single-cell resolution. The similarity between phenographs resulting from mutants within a given gene expression process (cf. each panel in Figures 2.3 and 2.4) suggests that our reporter can be used to test the relatedness within and between pre-mRNA processing events. It is interesting that numerous nonessential translation

mutants were analyzed, no appreciable effect on reporter expression was detected, suggesting no translation bias between the spliced and unspliced reporter RNAs (Figure 2.6). It is important to note that the red- and green-shifts can be a result of changes in one or both signals, as shown in Figure 2.5

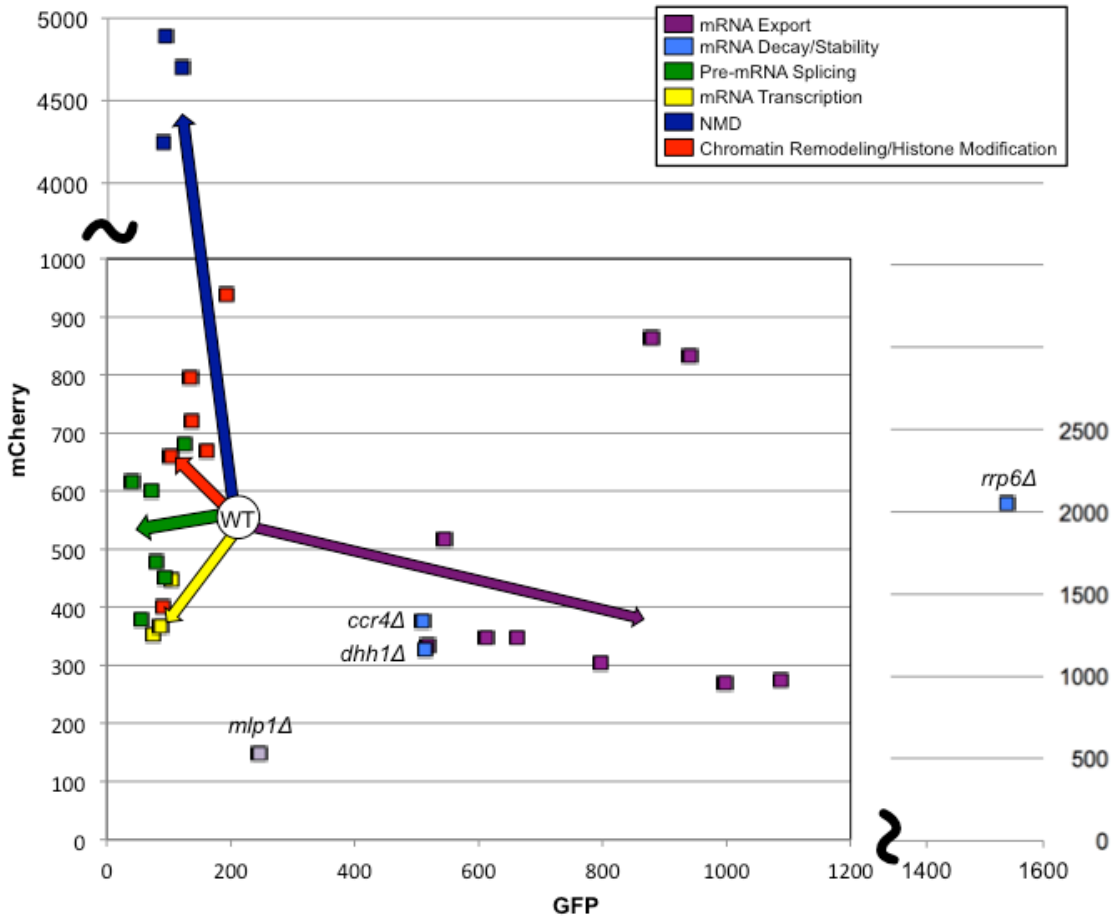


Figure 2.5: Comparison of total fluorescence readings from representative deletion mutants.

To describe the trend of the changes in both GFP and mCherry fluorescent signals, we plotted a number of mutants from Figures 2.3 and 2.4 based on the mean mCherry and GFP value in arbitrary fluorescence units for the 21,000 cells analyzed by flow cytometry. Note the scale of the inset graph is different that the scale for the larger graph, with the breaks in axes noted with a tilde. The arrows describe a vector for the average fluorescence change in the corresponding mutant group, with the arrowhead representing the actual average value for the mutant group.

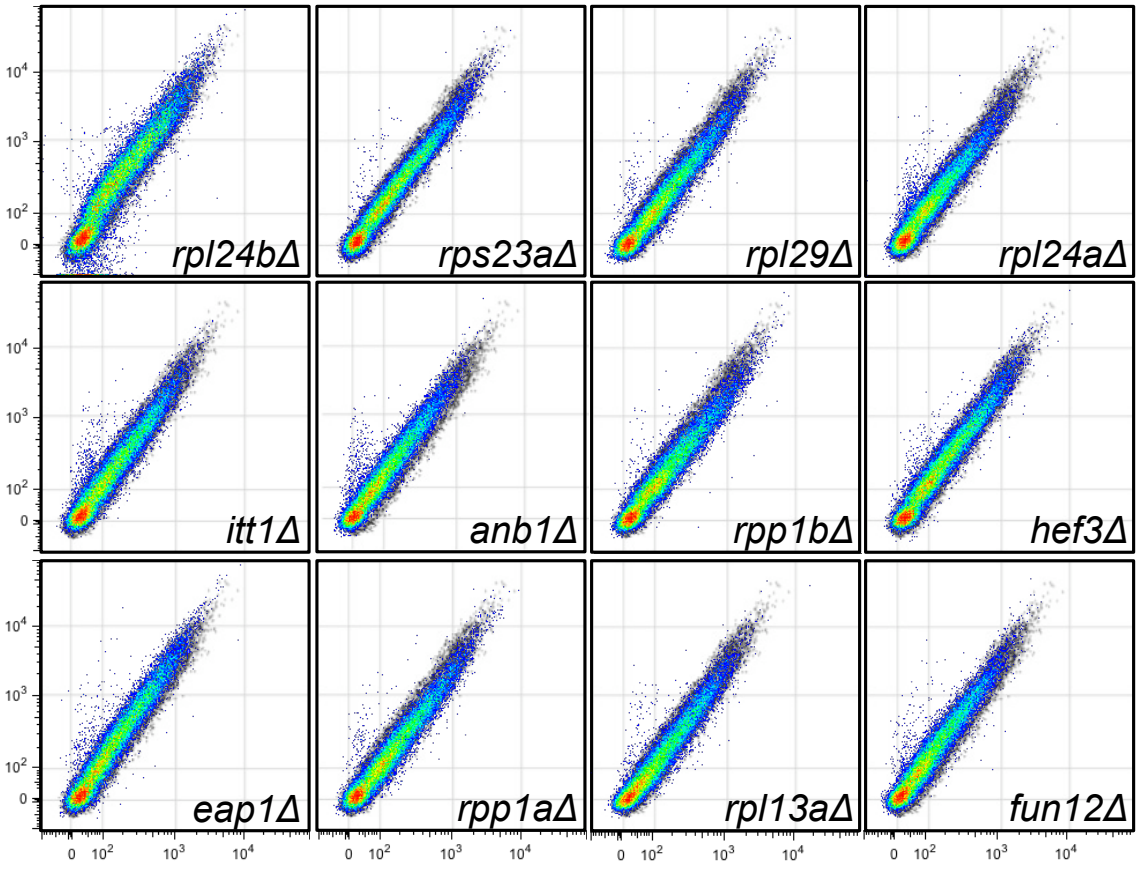


Figure 2.6: Nonessential translation-related deletion mutants have a minimal effect on reporter expression.

As in Figures 2.3 and 2.4, flow cytometry data from the indicated nonessential mutants were individually plotted and overlaid onto a grayscale dataset from wild-type cells. GFP and mCherry fluorescence are shown on the x- and y-axis respectively.

2.3.3 Optimization and utilization of a copper-inducible reporter

We also explored using our reporter with conditional alleles of essential genes involved in gene expression. Since most temperature- or cold-sensitive conditional alleles exhibit rapid response to changed temperature, the existing pools of extremely long-lived fluorescent proteins (Kitsera et al. 2007) would mask changes resulting from the mutation due to the relatively short time spent at the non-permissive temperature. To circumvent this, and to measure *de novo* signal after temperature shift, I placed the reporter described in Figure 2.1A under the control of the *CUP1* promoter, which is rapidly and highly activated upon addition of copper (Peña et al. 1998). Induction with 1 mM $CuSO_4$ resulted in significant fluorescence within one hour, reaching peak levels within 3-4 hours (Figure 2.7A). For completeness I followed the time course into stationary phase (6+ hours) and did not see any appreciable differences in reporter expression, as compared to the 4-hour time point. During optimization efforts I found that induction efficiency was highly dependent on growth phase, with the highest and most consistent levels produced when cells were induced between OD 0.1-0.2 (A_{600}, data not shown). Additionally, I have demonstrated that data from cells grown in standard culture tubes, 96-well, and 384-well plates are comparable using the *CUP1*-Reporter (Figure 2.8), simplifying the nature of high-throughput/high content experiments.

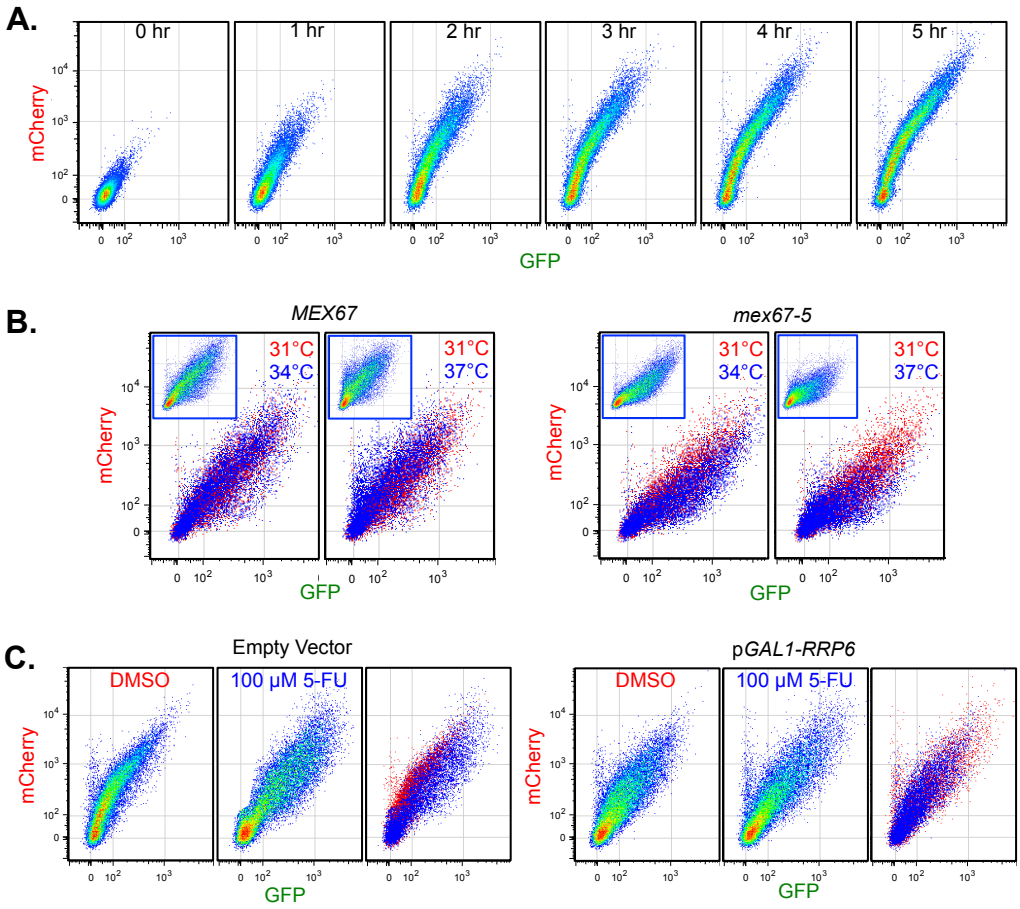


Figure 2.7: Exploring conditional alleles and drug treatment with a copper-inducible reporter.

A time course (A) of *CUP1*-Reporter induction is shown. Wild-type cells harboring a *CUP1*-Reporter were grown to early logarithmic phase and induced with 1 mM $CuSO_4$. Samples were taken at the indicated time points. Yeast with wild-type *MEX67* or *mex67-5* and the *CUP*1-Reporter were grown at permissive (31°C) temperature, then shifted to permissive or nonpermissive (34°C or 37°C) temperature with 1 mM $CuSO_4$ (to induce reporter) for 4 hours before analysis. Data for cells at permissive temperature are shown in the background in red, with data from cells at nonpermissive temperature in the foreground in blue (B). Pseudo color phenographs insets are shown for the corresponding nonpermissive conditions. Wild-type cells with *CUP1*-Reporter and an empty vector or *RRP6* overexpression plasmid were grown in minimal media with galactose. Cells were treated with DMSO (vehicle only control) or 100 µM fluorouracil (5-FU), induced with 1mM $CuSO_4$ and grown for 4 hours before analysis (C). The rightmost panel of each set is an overlay of the DMSO and 5-FU datasets for comparison.

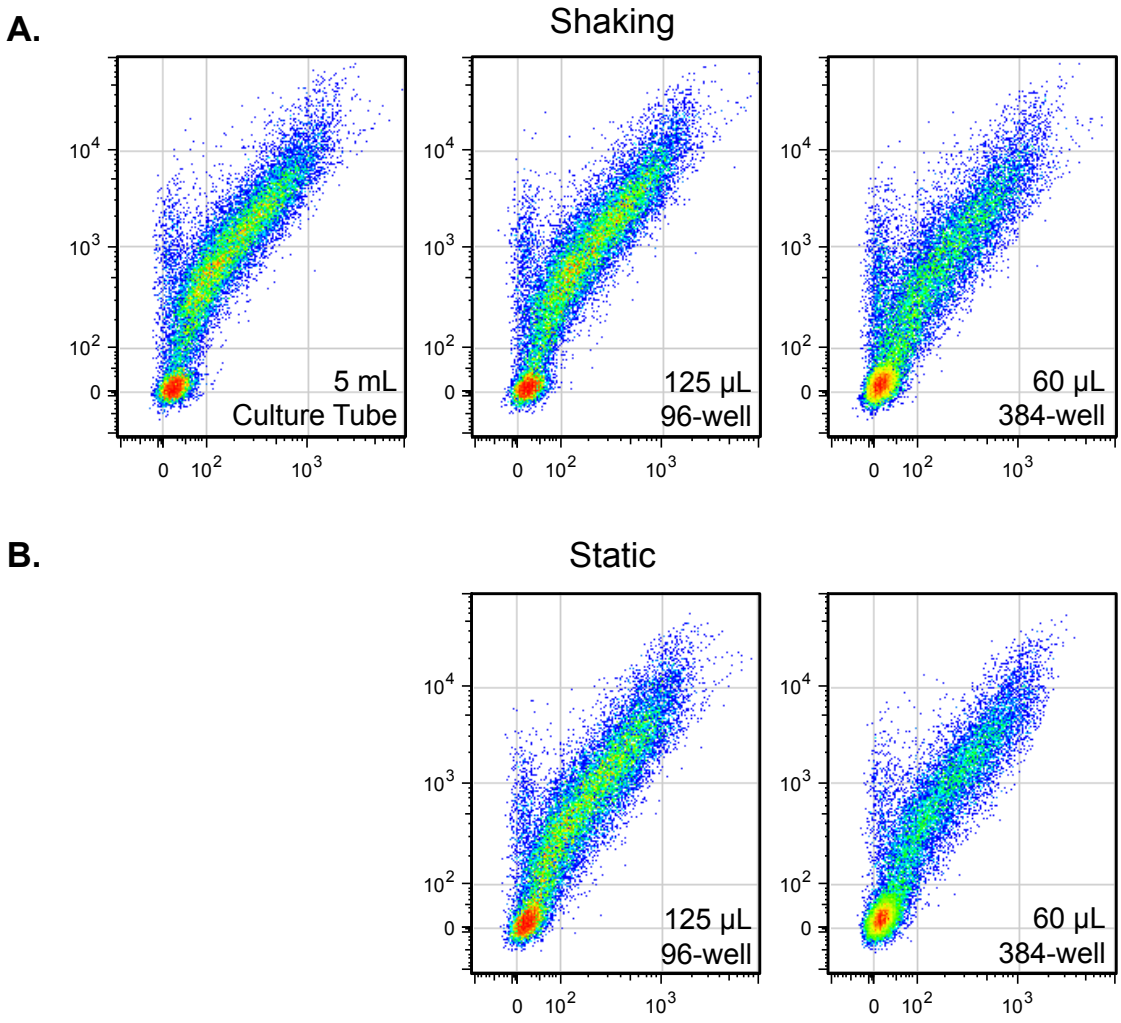


Figure 2.8: *CUP1*-Reporter expression in different culture conditions.

Wild-type cells harboring the *CUP1*-Reporter described in Figure 2.7 were induced with 1 mM $CuSO_4$, distributed into different growth vessels, and induced for 4 hours. Cells were then analyzed via flow cytometry (A). We also compared reporter expression in 96- and 384-well plates when grown without shaking (B).

Since we observed many mutants in the mRNA export pathway resulting in strong green-shifts, I analyzed the prototypical *mex67-5* allele of the essential Mex67p, homolog of the human TAP mRNA export factor (Segref et al. 1997). Yeast with the temperature sensitive *mex67-5* allele fail to export bulk mRNA at elevated temperature. I cultured *MEX67* (wild-type) or *mex67-5* yeast at permissive temperature (31°C) to early log phase, and then shifted to nonpermissive temperature (34˚C or 37°C) while adding 1 mM $CuSO_4$ to induce the *CUP1*-driven reporter. Yeast with wild-type *MEX67* displayed very little difference between permissive and nonpermissive temperature, while cells with *mex67-5* were green-shifted at nonpermissive temperature (Figure 2.7B). The green-shifted phenotype is consistent with the previously analyzed nonessential gene deletions in this process (Figure 2.4A-C), yet I observed more cell-to-cell variability with the conditional allele, which is perhaps not surprising since one would expect a more uniform phenotype with a complete gene deletion. I attempted to analyze the strong temperature sensitive splicing allele *prp2-1* but we found that the cells were not able to express sufficient levels of the reporter at nonpermissive temperature during the experiment, most likely due to the rapid growth defects conferred by the *prp2-1* allele (Hartwell 1967).

We were also interested in determining the utility of the inducible reporter with chemical insults that impair a specific gene expression process. Drug treatment of cells is similar to the use of conditional alleles in that existing pools of mCherry and GFP would not significantly change during the relatively short

treatment, thus the inducible reporter would add great utility to these experiments. The nucleoside analog fluorouracil (5-FU), a widely used chemotherapeutic (Longley et al. 2003), has been shown to inhibit the 3'-5' exonuclease Rrp6p of the nuclear exosome (Fang et al. 2004; Kammler et al. 2008), although the exact mechanism of inhibition is unknown. As expected, treatment with 100 μM 5-FU while inducing the *CUP1*-Reporter resulted in a green-shift, compared to a DMSO only control (Figure 2.7C), and mimicked the deletion of *RRP6* (Figure 2.4H). Additionally, I show that overexpression of *RRP6* is able to reverse the effect of 5-FU on reporter expression (Figure 2.7C). Together, the *mex67-5* and 5-FU data demonstrate the usefulness of the inducible *CUP1*-Reporter in analyzing essential genes and chemical perturbations.

2.3.4 Description of our analysis pipeline

The nature of our reporter also allows for rapid high-throughput flow cytometry analysis, permitting one to acquire data from hundreds of unique strains or conditions within a matter of hours, while collecting tens of thousands of data points (events or cells) for each sample. With the significant amount of data our reporter system generates, we sought out ways to compare reporter expression in many different samples in an unbiased manner. Since this type of analysis platform was lacking for flow cytometry data, we developed a statistical workflow that vectorizes the data, describing the phenograph shape by reporting

the distribution of fluorescent events in the two-dimensional phenographs. These vectorized data are then compared to each other using a standard hierarchal clustering algorithm. We named our analysis pipeline “CHUBACA” (Comparative Hierarchy Using Binning And Clustering Analysis).

A graphical representation of CHUBACA is shown in Figure 2.9. To begin, we used FlowJo to process FCS files. Scale values for GFP and mCherry fluorescence for all events were exported into a text file, resulting in an individual file for each sample containing a column for each value and a row for each event. These files were then processed within SAS using our script to bin the events based on their position within the 2D plot (demonstration of binning order and pattern is shown in Figure 2.10). To adjust the scale of each axis, the minimum and maximum values can be independently adjusted according to the needs of the user. Additionally, the resolution of each axis can be tuned by adjusting the number of evenly spaced bins. We found that empirical determination with technical and/or biological replicates is essential for determining the appropriate resolution (bin number). The binning results for all samples were exported as a single, tab-delimited text file. The first row and column of the resulting file contained the bin number and sample names respectively. Each row contains the number of events contained within each of the bins. The data were then clustered hierarchically (using Cluster) to create the CDT and GTR files from which Java TreeView generates the dendrogram files.

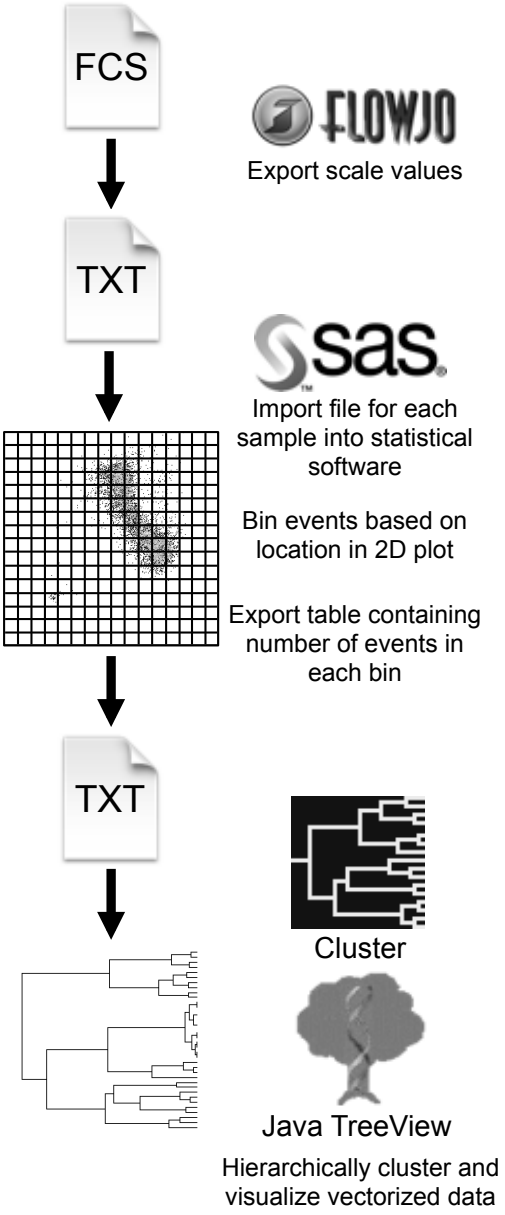


Figure 2.9: Graphical representation of the CHUBACA analysis pipeline.

Briefly, standard FCS files are processed in FlowJo and scale values for two parameters are exported into a text file (TXT). Individual files for each sample are imported into SAS, binned, and a single TXT file is created describing the number of evens in each bin for all samples. These data are then hierarchically clustered and analyzed/viewed using Cluster and Java TreeView. Further details are described in the Materials and Methods and Section 2.2.5.

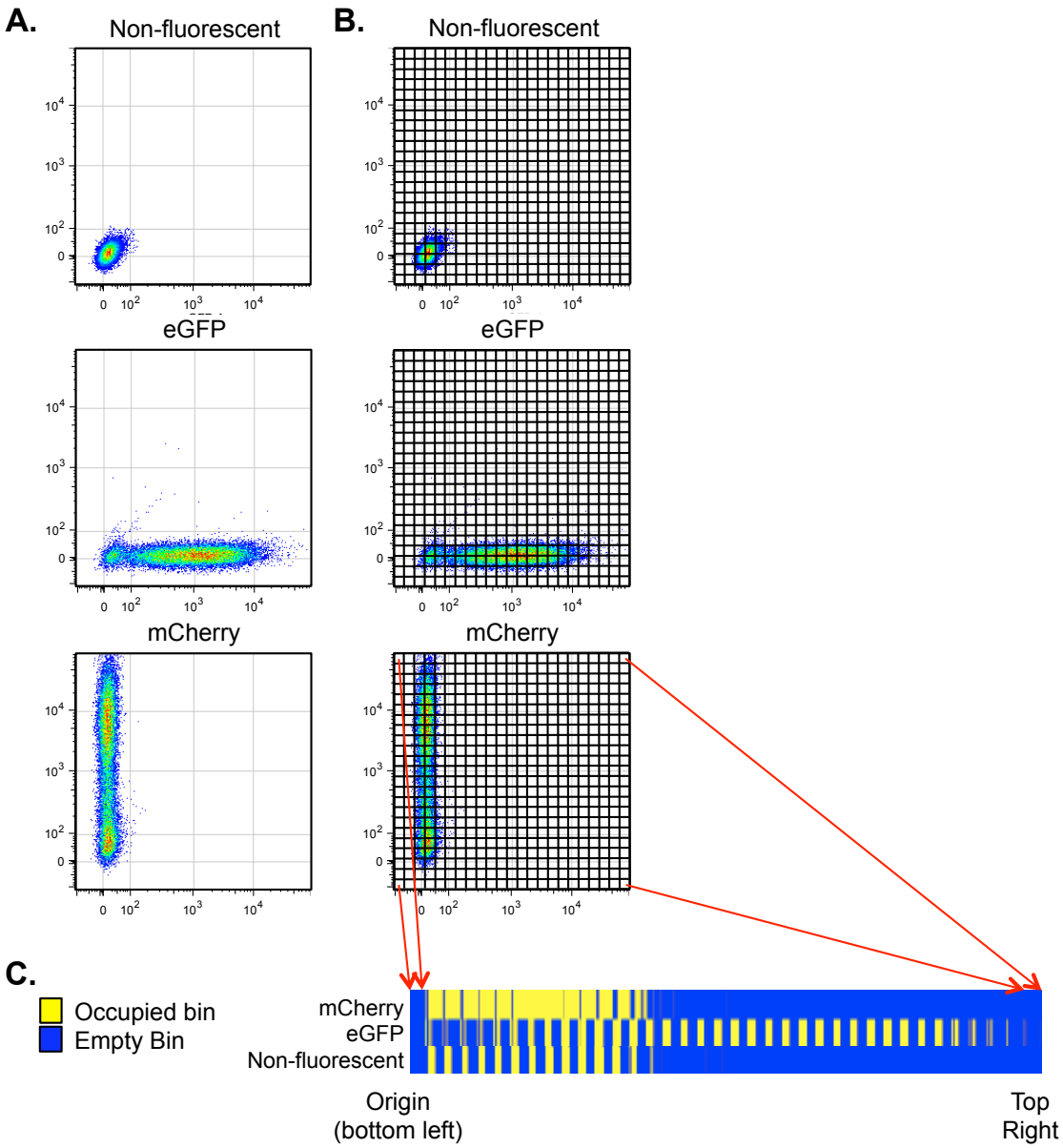


Figure 2.10: Description of binning and clustering process.

Flow cytometry data (A) is exported from FlowJo and imported into SAS. An algorithm then sorts the individual cells within the phenograph into *n* x *n* bins (B) and reports the number of cells in each bins. The bins are ordered starting from the origin (see arrows) and up the y-axis. It shifts rightward to the next column, goes up again etc., ending with the top right bin. These data are then exported as a TXT file, imported into Cluster for hierarchical clustering. The occupancy patterns are shown for our fluorescence controls (C) to demonstrate the binning pattern.

2.3.5 Demonstration of bin density importance in CHUBACA

As stated earlier, CHUBACA allows for fine-tuning of axis (parameter) settings to allow for a wide range of values as well as an adjustment of resolution. To demonstrate the utility of optimizing the number of bins in the analysis, I used data from dozens of deletion mutants containing our gene expression reporter. This dataset contained 68 samples, including a number of biological replicates and controls. 20,000 events were binned for each sample and the results were hierarchically clustered. In Figure 2.11A, I demonstrate the effect of iteratively increasing the number of bins applied to the 2D plot, resulting in 100, 400, 1600, or 6400 bins (10, 20, 40 or 80 bins per axis). When observing the clustering pattern, one can appreciate the influence of bin number on the resulting analysis by first comparing the clustering behavior of the fluorescence controls. For instance when I binned with a 10x10 grid, the mCherry-only controls cluster in a clade separate from nonfluorescent cells and wild-type samples with the reporter (Figure 2.11A, 1st panel). When I increased the bins to 20x20 (Figure 2.11A, 2nd panel), the mCherry samples join the nonfluorescent and wild-type samples and then cluster separately when further increased to 40x40 and 80x80 bins (Figures 2.11A, 3rd and 4th panels, respectively). With increasing bin number I also observed deeper and more distinctive clades. This may be expected as when the same data are separated into an increasing number of bins, the resulting data become more distinct between samples, essentially increasing the resolution of the analysis. Consistent with this trend is the decreasing correlation

coefficient within a specific clade. For example the correlation between the 6 nonfluorescent controls is 0.877 with 100 bins, decreasing to 0.719 with 6400 bins.

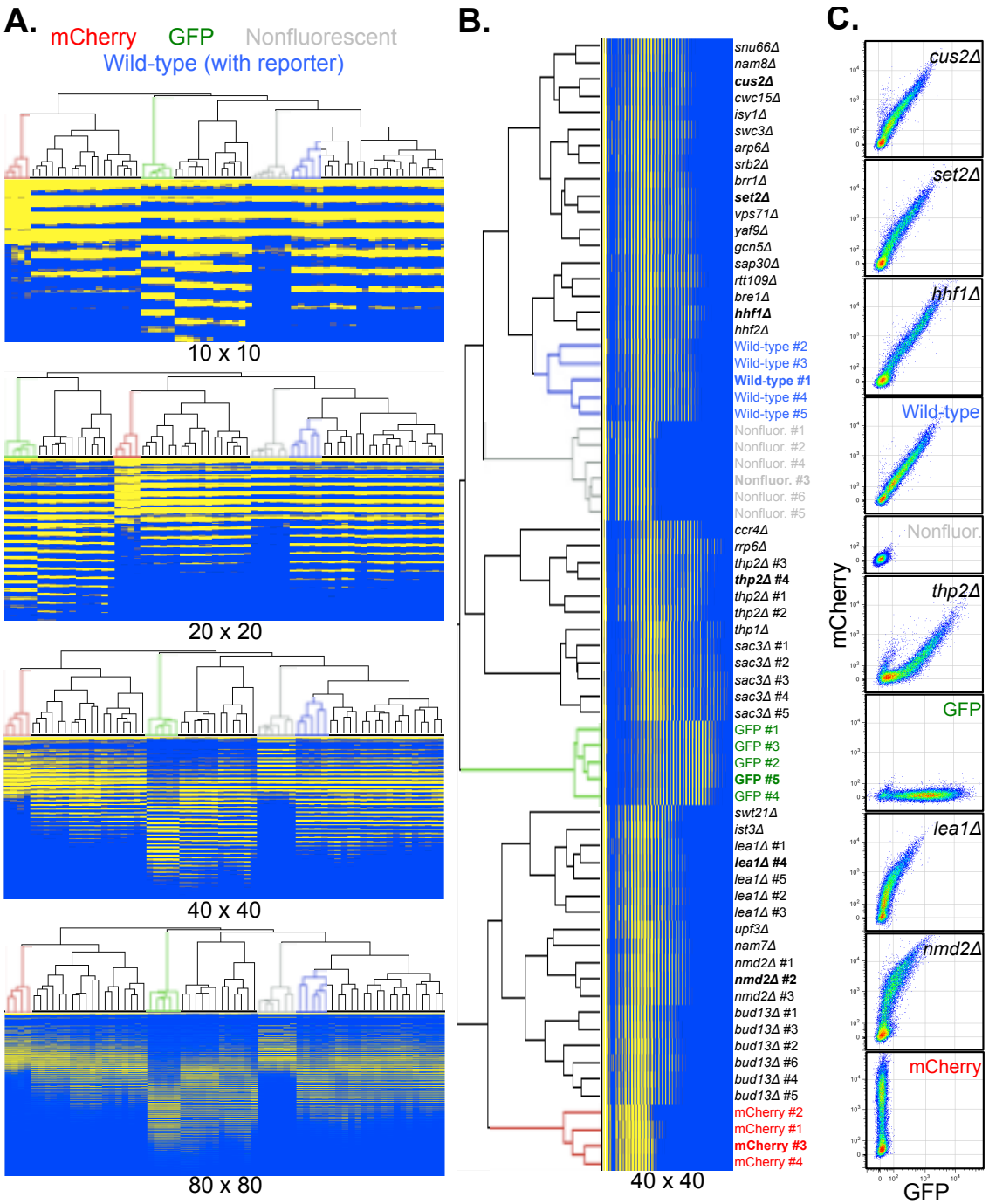


Figure 2.11: Fine-tuning and optimization of the binning algorithm.

(Figure 2.11 continued)

Dendrograms for the analysis of 68 yeast samples are shown (A). A legend for annotated controls is shown at the top and nodes/branches for the aforementioned groups are colored respectively. Below each dendrogram are the vectorized data indicating the degree of bin occupancy (yellow, occupied; blue, unoccupied). Bin number densities are indicated below each dendrogram and bin occupancy map. (B) Expanded, full annotation of the optimal 40x40 bin density dendrogram. The same colors used in (A) are used to highlight control samples. Bolded sample names are represented in (C). (C) Representative flow cytometry data from the indicated samples are shown in pseudocolor. 8K events are shown.

It is critical to note the importance of incorporating out-groups and biological replicates to empirically determine the optimal settings when binning data from a given dataset or experimental system. When identifying the axis range, I limited the range as much as possible within the data being generated from our system to provide the analysis adequate separation between events. In terms of bin number I monitored the clustering behavior of biological replicates and genetic or chemical perturbations that result in similar effects. For instance, I find that using 40 bins on each axis (Figure 2.11B) is ideal for our reporter assay since the non-wild-type sample *sap30Δ* clustered with wild-type samples when we increase the bin number beyond this point (Figure 2.11A, 4th panel). Another example is that of the biological replicates of *lea1Δ* which clustered together in Figure 2.11B and are separated and include related but not identical mutants when the resolution is increased beyond 1600 bins. A number of representative 2D phenographs from the minimized dataset are shown in Figure 2.11C.

2.3.6 Multi-parametric analysis in a comparison of T cell activation

Experiments utilizing flow cytometry often examine multiple parameters simultaneously. To demonstrate that our analysis pipeline is amenable to typical flow cytometry samples such as those derived from human blood cells, I analyzed a publically available dataset (see Materials and Methods Section 2.2.4). The 36 samples analyzed involve the titration and optimization of ligands leading to human T-cell activation. Over a 10-minute time course, the

researchers queried two different anti-CD3 antibodies (clone OKT3 or UCHT1) at high or low concentration in the presence or absence of anti-CD28 antibodies (see Materials and Methods Section 2.2.4 for further detail Landrigan et al. 2011). Within this time course, a transient peak in both phosphorylated SLP76/S6 (pSLP76) and phosphorylated CD247/TCRζ (pCD247) is evident at the 3-minute time point. After this time point levels decrease, reaching near baseline levels by 10 or 20 minutes, agreeing with previous studies showing similar SLP76 phosphorylation kinetics in Jurkat cells (Houtman et al. 2005). In my analysis, I binned the data for multiple 2D comparisons beyond the primary parameters of interest (which were phosphorylated SLP76/S6 (pSLP76) and phosphorylated CD247/TCRζ (pCD247)). In addition to the pSLP76/pCD247 comparison, I also binned the naïve CD45+ T cells by FSC/SSC and CD45RA/CD45RO, which were used to gate living and naïve cells respectively. Since there was inevitably less variability among the parameters that were used in gating, I expected much less distinction between samples when looking at these comparisons.

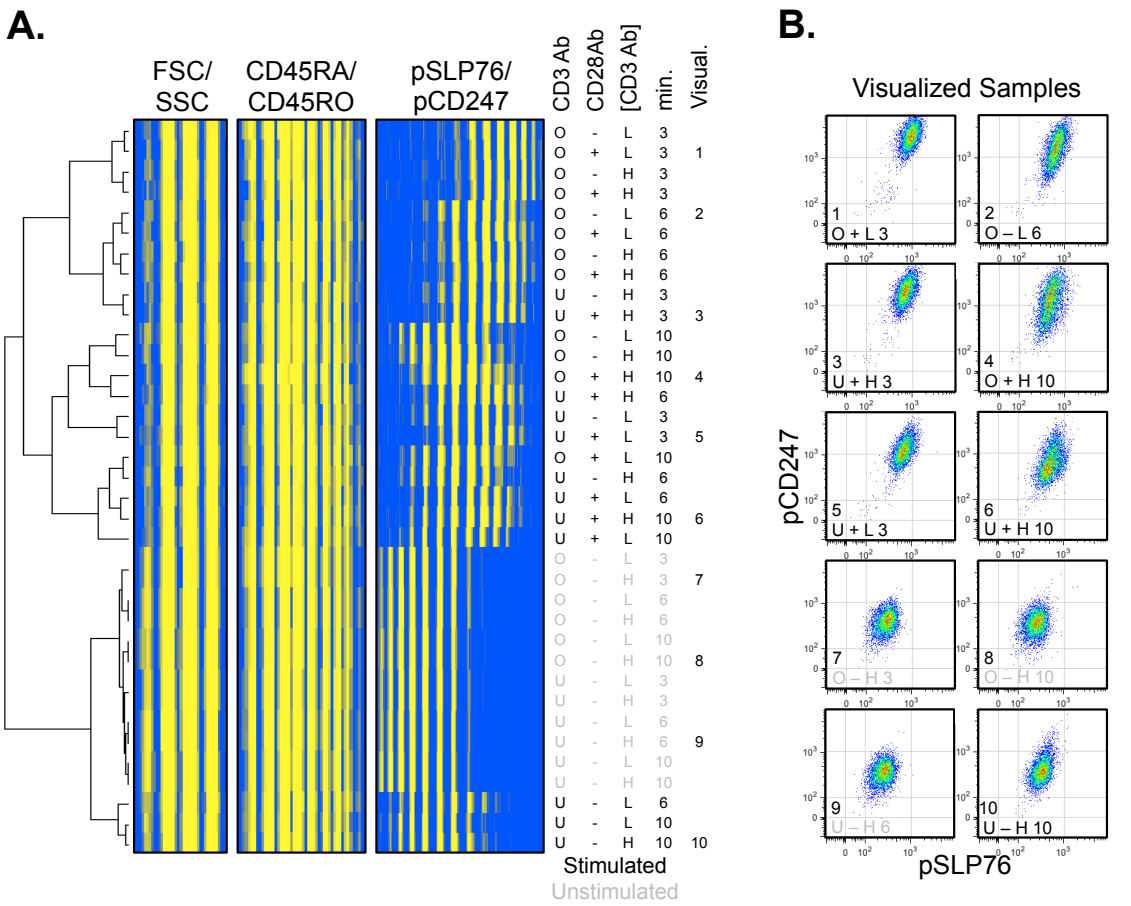


Figure 2.12: Multiparameter binning and clustering analysis of a T cell stimulation dataset.

Binned, six-parameter flow cytometry data from 36 samples of naïve T cells (gated by FSC-A/SSC-A, FSC-A/FSC-H, and CD45RA/CD45RO for live, single, naïve cells respectively) (nodes/rows) were combined and clustered hierarchically (A). The three blue/yellow bin arrays were offset for easier visualization and depict occupied (yellow) and unoccupied (blue) bins within the 2D plots labeled above (bins which were empty in all samples were removed from the analysis). Samples are labeled to the right with the type of CD3 antibody (U, UCHT1; O, OKT3); absence (-) or presence (+) of 10 µg/mL CD28 antibody; low (0.3 µg/mL, L) or high (0.8 µg/mL, H) concentration of CD3 antibody; stimulation duration in minutes; and a reference number if visualized in panel B. Samples labeled in grey are unstimulated vehicle-only controls. Eight thousand events from samples labeled with a number in the “Visual.” column of panel A are shown in pseudocolor (B), reflecting phosphorylation levels of SLP76 and CD247 on the x- and y-axis respectively. Internal labels correlate with the visualized sample number and treatment conditions.

Operationally, each of the three binning analyses was performed in series and individual results were combined prior to clustering, ensuring that all 6 parameters were utilized in the comparison. In Figure 2.12 I illustrate the results of the six-dimensional binning and clustering analysis of naïve T-cells using 1600 bins per analysis. Samples are labeled to the right of the dendrogram and 10 specific samples are labeled arbitrary (Samples 1-10) in the far right column to aid in discussion. The 36 samples separated into two main groups, the top group possessing higher levels of pSLP76 and pCD247 and the lower group possessing lower levels. When examining the binning patterns, it is clear that there is much more variation within the pSLP76/pCD247 plot (right) compared to FSC/SSC and CD45RA/CD45RO (left and middle), agreeing with our earlier prediction. Twelve unstimulated controls cluster very tightly together near the bottom of the panel. In agreement with their time-course data (Landrigan et al. 2011), we observe many of the shorter (3 minutes) stimulation time points, at which time pSLP76 and pCD247 levels peak, cluster together at the top of the dendrogram.

To illustrate how the flow cytometry data appear, I display pSLP76 and pCD247 levels for 10 of the 36 samples in Figure 2.12B. When looking at samples with high pSLP76 and pCD247 levels, I noticed the shorter time points resulted in more significant shifts from the unstimulated state (top of Figure 2.12B). This is not only seen in the pSLP76/pCD247 plots, but also in the binning pattern shown in Figure 2.12A. More bin occupancy is seen towards the right,

relating to the top-right section of the plots, compared to the other samples. Additionally, the presence of anti-CD28 does not significantly affect the levels of pSLP76 and pCD247, which also agrees with the published analysis (Landrigan et al. 2011). I also observed that the higher concentration of anti-CD3 typically resulted in slightly higher levels of pSLP76 and pCD247, which is likely why the authors chose to use a higher concentration in their subsequent work.

2.3.7 Reporter is amenable to high-throughput analysis

The flow cytometry datasets analyzed thus far did not take full advantage of the scalability of our analysis pipeline, CHUBACA. To demonstrate the high-throughput nature and reproducibility of our reporter and analysis pipeline, as well as to query all nonessential yeast genes for their role(s) in gene expression, I completed a high-throughput screen of 5103 viable strains in the yeast haploid deletion collection (Winzeler et al. 1999). Individual transformants were obtained for 5020 strains (98%), and were individually analyzed by high-throughput flow cytometry. Data conforming to our criteria were obtained for 4967 mutants (97% of viable strains, 99% of transformed strains). To first assay the reproducibility of our assay across the multiple days this experiment took place, we observed the clustering behavior of data from biological replicates assayed on different days. In Figure 2.13A I show the clustering behavior of the deletion collection data with the biological replicates added. Figures 2.13B and 2.13C enlarge areas of interest and reveal the tight clustering of fluorescent controls, as well as

previously described gene expression mutants such as *sac3*Δ, *lea1*Δ, and *bud13*Δ. The replicates clustered very tightly together and near genes of similar function within this large dataset, further demonstrating the reproducibility and robustness of our assay.

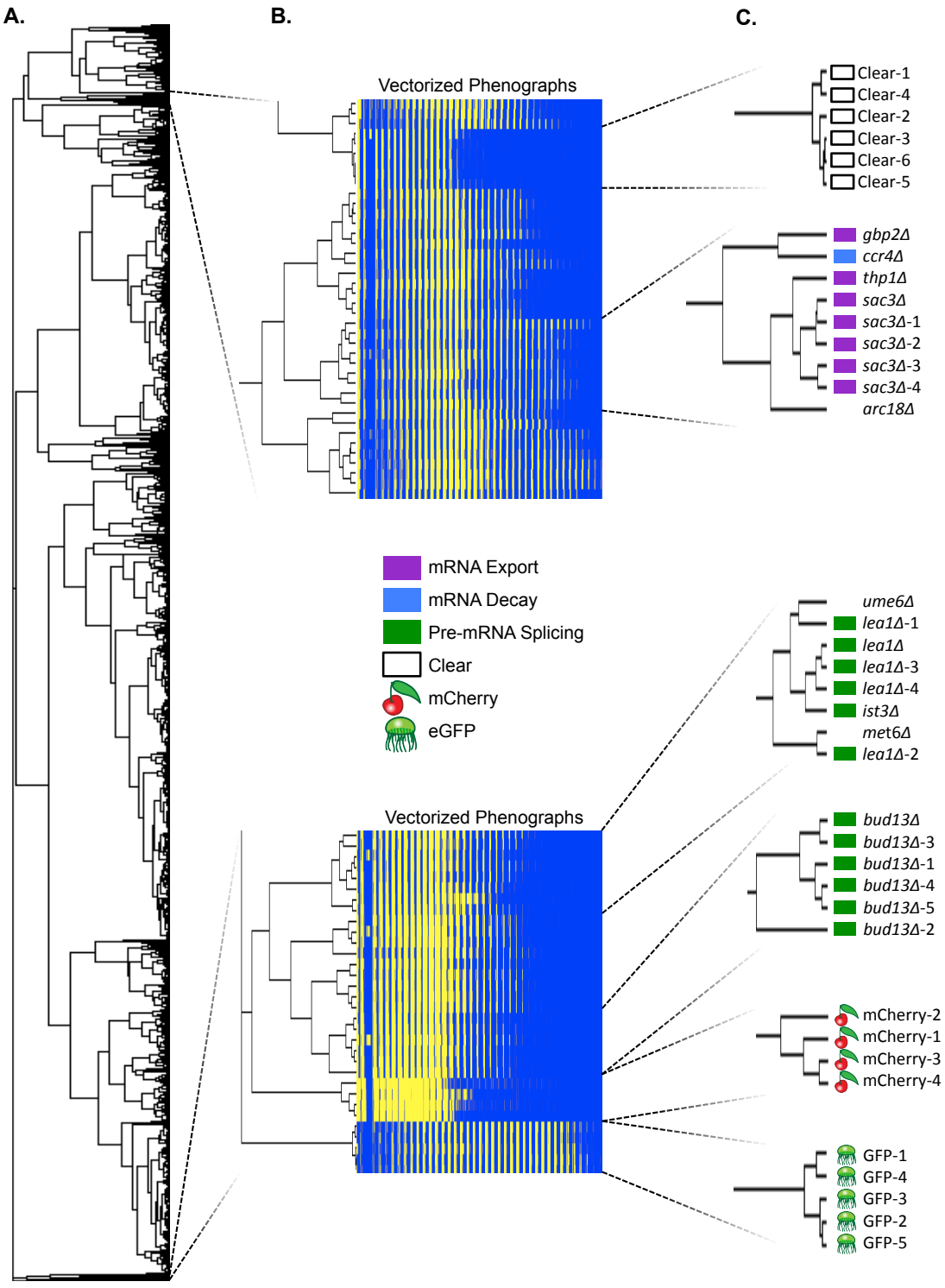


Figure 2.13: Robustness and reproducibility of the reporter assay demonstrated with the deletion collection screen data.

(Figure 2.13 continued)

Phenograph data for fluorescent controls and replicates of telltale deletion mutants were incorporated into the deletion collection dataset and clustered (A). The control data were taken from random days of large-scale analysis and serve as day-to-day controls to monitor experimental reproducibility. Phenograph data for the deletion mutants and control samples were binned and clustered using complete linkage and an uncentered correlation similarity metric and the complete dendrogram is shown. Clusters of interest were enlarged and the binning pattern for several mutants is shown (B). The bottom panel (mCherry and GFP) clearly demonstrates the relationship between phenograph and binning pattern. Bins are ordered from the origin to the top right, increasing in y value till the top, then shifting to the next x bin, etc. (explained visual in Figure 2.10). Dash numbers represent biological duplicates of controls sampled on different days and are shown in the enlargements (C). Nodes without a dash are the yeast strains that were analyzed within the original screen. Nodes are color-coded by function or sample indicated in the legend.

2.3.8 Clustering behavior of the haploid deletion collection

Since most of the strains in the deletion collection are not related to gene expression, it was reasonable to expect that the vast majority of strains would be highly similar to wild-type (Figures 2.1C and 2.3G). I observed that 87% of the deletion mutants clustered within a very large clade (Figure 2.14A bottom). The limited variation observed between the mutants in this group likely describes biological noise or mild indirect effects on gene expression.

Within a small clade near the top of the dendrogram (Figure 2.14A), a group of yeast strains (*n*=16) was detected that was enriched with mutants lacking mRNA export factors (n=*9*; Figure 2.14B). Note the tight clustering of *sac3*Δ and *thp1*Δ (Figure 2.14B), which was expected due to their nearly indistinguishable phenographs (Figure 2.4A) and their roles in the TREX-2 complex. Moreover, there were other export factors such as additional TREX-2 components (*sus1*Δ and *sem1*Δ), THO complex (*thp2*Δ), nuclear pore components (*mlp1*Δ and *nup120*Δ) and the poly(A) RNA binding factor *gbp2*Δ. Also within this clade were mutants lacking genes that influence mRNA decay or stability including *dhh1*Δ (decapping), *ccr4*Δ (deadenylation) and *rrp6*Δ (nuclear exosome).

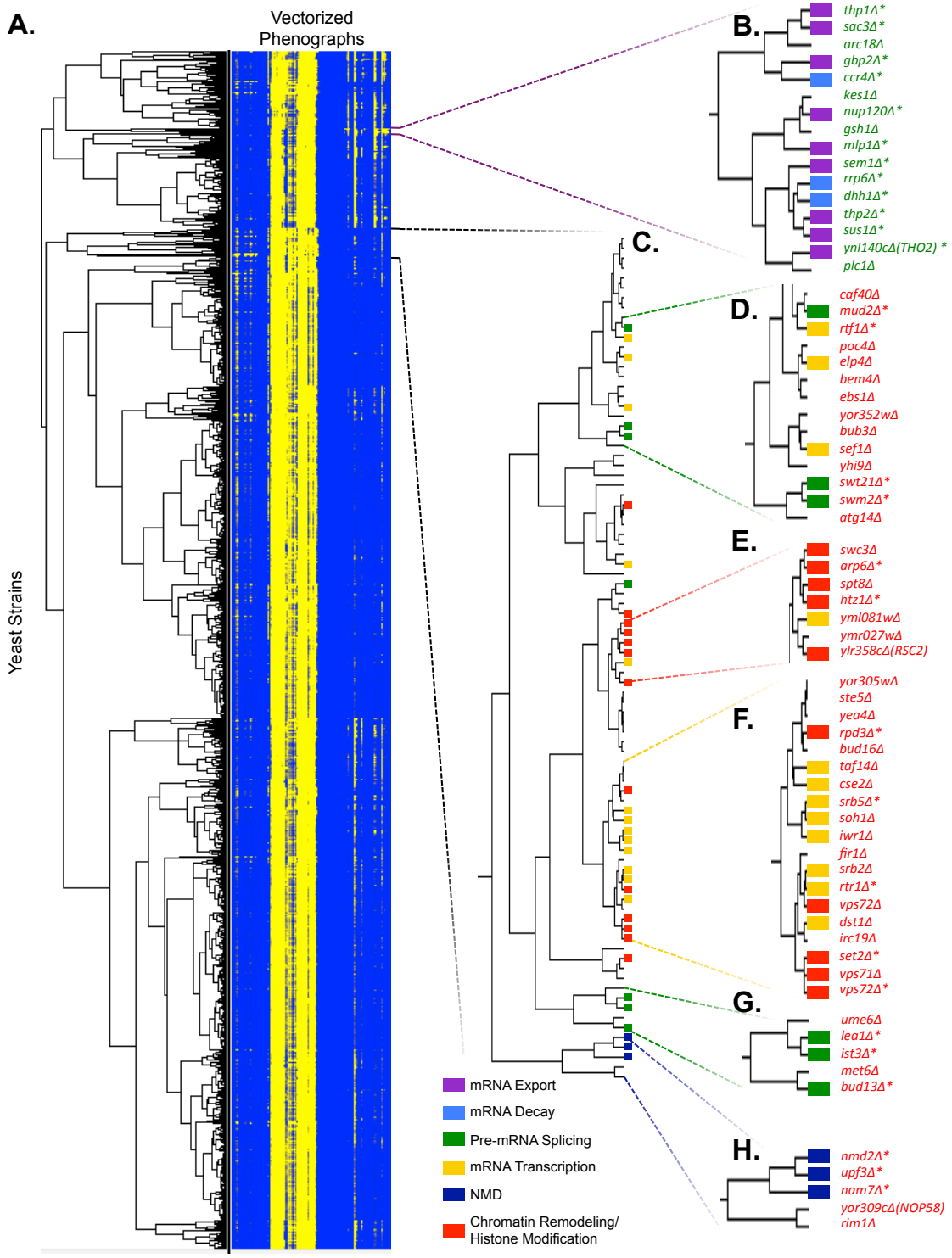


Figure 2.14: An unbiased comparison of phenographs from the deletion collection strains reveals clustering of related genes.

(Figure 2.14 continued)

Clustered, vectorized phenograph data for the 4967 deletion collection strains as calculated by Cluster (using complete linkage/uncentered correlation) and visualized by Java TreeView (A). The 1600 bins (columns) are ordered from the origin of the phenograph to the top right of the phenograph and correspond to the same bin across all strains analyzed (rows). Yellow shading indicates the level of occupancy in the given bin. A number of green-shifted deletion mutations represent mRNA export and RNA decay mutants and cluster strongly (B). The gene deleted is indicated and functional classes of genes are color-coded according to the scale at the bottom. A representative phenograph for a green-shifted mutant overlaid on a grayscale wild-type is shown. Asterisks indicate strains with phenographs in Figure 2.4. Genes in parentheses are partially overlapped by the deleted dubious or unverified ORF. A midlevel view (C) of the clustering behavior of a large red-shifted clade containing a number of pre-mRNA splicing mutants in shown. Clusters of specific red-shifted gene expression deletion mutants (D-H) are shown. Branch lengths are kept to scale in elements to demonstrate levels of similarity within and between clades. For reference, asterisks indicate strains with phenographs in Figures 2.3 and 2.4.

Next, I investigated the clustering behavior of pre-mRNA splicing mutants and other red-shifted mutants within the larger red-shifted clade (*n*=85; Figure 2.14C). The smaller groups within this clade are each enriched with factors within certain gene expression processes such as pre-mRNA splicing, transcription, chromatin remodeling, histone modification and NMD (Figures 2.14D-H). The pre-mRNA splicing mutants clustered into two slightly offset clades within this group (Figures 2.14D and 2.14G). This suggests that our reporter is differentially sensitive to pre-mRNA splicing mutations and quantitatively illustrates the difference in the phenographs we reported in Figures 2.3A and 2.3B.

Within this red-shifted group, there was enrichment of chromatin remodeling and mRNA transcription mutants (Figures 2.14B, 2.14E, 2.14F). Adding to the confidence that the clustering of the reporter data correlates with functionally related genes, the chromatin remodeling mutants in this clade contain components of the SWR1 complex (*swc3*Δ and *arp6*Δ) and also the H2AZ histone variant (*HTZ1*) on which it operates (Mizuguchi et al. 2004). Similarly, the mRNA transcription clade contained deletions of constituents of the mediator complex (*cse2*Δ, *srb5*Δ, *soh1*Δ and *srb2*Δ; (Guglielmi et al. 2004)). The most tightly clustering group within this red-shifted clade contained the NMD mutants (Figure 2.14H), which was expected due to their distinct phenograph shape (Figures 2.3C). The remarkable degree to which the transcription, histone modification, and chromatin remodeling mutants cluster among pre-mRNA

splicing mutants (Figure 2.14C) is not surprising, given that most introns are believe to be removed cotranscriptionally (Brugiolo et al. 2013).

2.4 Discussion

When compared to existing yeast pre-mRNA splicing reporters for budding yeast, the reporter described here expresses well from a low-copy vector, is able to simultaneously measure both pre-mRNA and mRNA levels, and is amenable to rapid, sensitive gene expression analysis in single cells via flow cytometry without the need to prepare extracts from cells. It is important to note that the red- and green-shifts are not only due to changes in splicing efficiency, per se, but can also be a result of the localization and relative stability of either the pre-mRNA or mRNA, as exemplified by mRNA export mutants (Figures 2.4A-C and 2.7C) and mRNA degradation and stability mutants (Figures 2.3C and 2.4F-H). Please refer to Illustration 2.1 for visual representation of the gene expression processes that influence our reporter's behavior and their effects. The use of flow cytometry to analyze reporter expression also permits simultaneous observation of cell volume and complexity using the forward and side scatter measurements, respectively. These features enable us to assay the noise in eukaryotic gene expression among a population of cells in a unique way (Tang et al. 2011; Waks and Silver 2010), as well as simultaneously examining effects on yeast morphology with forward and side scatter values.

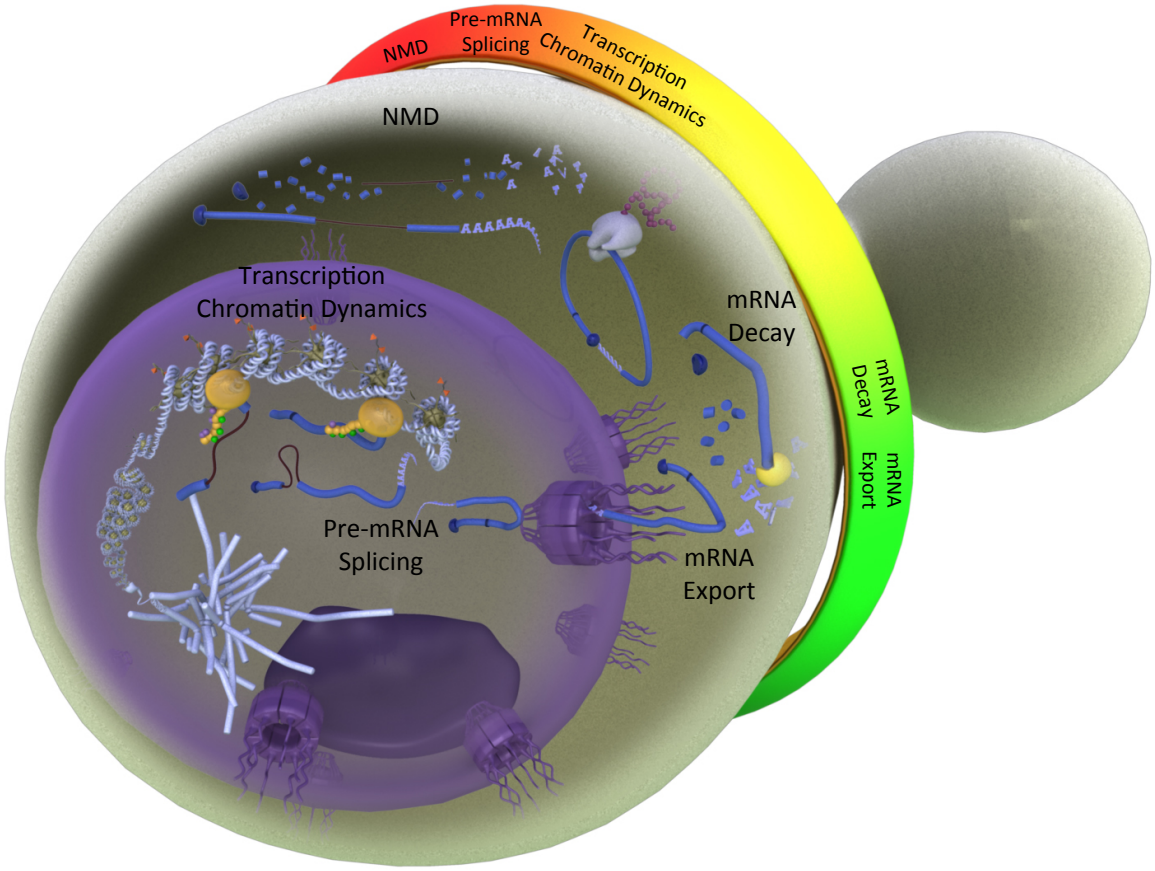


Illustration 2.1: Demonstration of gene expression reporter using a yeast "mood ring" model

The gene expression processes that our gene expression reporter is sensitive to are labeled within the cartoon yeast cell. The processes are then labeled on the colored "mood ring" to demonstrate the resulting color on a red-to-green spectrum. Wild-type cells would be in the yellow area. This visualization was made possible thanks to the creativity and skill of Mariana Grenadier, Computer Illustrator at the College of Natural Sciences.

In the design of our reporter, one goal was to maximize its utility compared to previously described gene expression reporter. Since our readout was single-cell in nature, we also desired to use a stable low-copy centromeric plasmid to potentially limit the amount of cell-to-cell variability due to copy number variation, as well as to express our reporter at more natural levels. The single-cell aspect of our reporter also permits the investigator to use our reporter and fluorescence activated cell sorting to identify clones of interest in a rapid and unbiased manner when used in conjunction with conventional genetic screens. This assay should be feasible and cost effective for researchers at research institutions that have access to core facilities with flow cytometry capabilities. Additionally, with the advent of many small-footprint tabletop flow cytometers, it is also within reach for many investigators to complete the flow cytometry analysis within their own laboratory. Also, by avoiding the use of endogenous reporter genes, and by cloning the reporter into plasmids with various auxotrophic and dominant selection markers, we have virtually eliminated genotypic restrictions. Moreover, I expanded the application of our reporter by developing the copper-inducible *CUP1*-reporter, and demonstrated its usefulness with conditional mutants of essential genes (*mex67-5*), as well as with treatment of 5-FU. During development of the *CUP1*-reporter, we also examined reporter expression in the same mutants as in Figures 2.3 and 2.4 and compared expression to the *TDH3*-driven reporter saw patterns and shifts of similar magnitude (data not shown).

In our survey of nearly 5000 knockout strains, our novel gene expression reporter proved to be highly sensitive and specific for numerous gene expression-related processes. This dataset, as well as binned data for the biological replicates and controls in Figure 2.13, are available with our published work (Sorenson and Stevens 2014). This will allow the community to perform similar analyses in additional strains or under difference conditions and directly compare their data to learn more about their genes or alleles of interest. Additionally, researchers can reanalyze these data and look for their gene of interest to identify knockout mutants with similar effects on reporter expression, as well as refer to a list of red- and green-shifted mutants (Appendix II). Our ability to directly compare the expression of the reporter between strains by means of the vectorized phenograph, and the fact that genes participating in related functions behave similarly by these analyses makes this a useful tool for future studies. Within our clustering analyses of the deletion collection, we noticed that mutations predicted to lead to a completely defective, non-essential pathway (e.g. NMD) cluster tightly with each other. This is in contrast to mutations of components of essential machineries (e.g. pre-mRNA splicing), which cluster less-tightly and result in decreased efficiency, but not a complete loss. Lastly, we observed that mutants containing deletions of dubious or unverified ORFs regularly clustered near deletions of verified ORFs with which they overlapped, such as *ynl140c*Δ and *tho2*Δ, *ydl094c*Δ and *pmt5*Δ, and *ydr290w*Δ and *rtt103*Δ respectively.

I also demonstrated that our reporter assay was robust and reproducible. When I included reporter data from biological replicates of wild-type cells into the deletion collection data, we observe that they cluster with deletion mutants unrelated to gene expression (cf. Figures 2.2C and 2.3G) that have wild-type-like phenographs. The reproducible behavior of mutants with a more distinct phenograph (Figure 2.13) demonstrates the robustness of our assay and analysis pipeline.

A very powerful aspect of our reporter is that it is sensitive, yet uniquely specific, to defects in many different gene expression events, including the processes involved in both biogenesis of mRNAs as well as their degradation. This allows for a researcher to simultaneously assay the effects of a wide variety of stimuli on many processes. We have found that our analysis pipeline using our binning and clustering technique, as well as our telltale phenograph patterns (Figures 2.3 and 2.4), quickly identifies potential pathways to explore further with orthogonal assays, such as the RT-qPCR we present in Figure 2.2B. Additionally, multiplexing of reporters can also be a way to screen for effects on multiple reporters simultaneously (Chen et al. 2012; Kulak and Lum 2013), as long as their signals do not overlap or disturb each other. But as with all other reporter technologies, it is still critical to ascertain the effect on the original or endogenous targets.

Compared to traditional microscopy, which is prone to subjective biases, flow cytometry is relatively automated and can provide quantitative,

multiparameter output for thousands to millions of cells per sample. Thus flow cytometry has become a trusted platform in biomedical research and in clinical settings. There is an increasing need for analyses that can identify specific cell populations within a complex sample and identify differences and similarities between samples across multiple parameters. Although there are many robust and sensitive platforms that identify specific populations of interest within a sample, our platform, CHUBACA, compares overall populations across any number of parameters by comparing the shape of one or more 2D plots. We demonstrated that CHUBACA not only works for our two-color yeast reporter system (Sorenson and Stevens 2014), but that it is appropriate for conventional flow cytometry data of primary samples.

We have found that in the wide range of analyses performed, the binning settings work well between experiments once empirically determined for a specific experimental system. Additionally, if flow cytometry settings are carefully calibrated between experiments, the data from separate experiments can be meaningfully compared. As discussed above when describing the effect of changing the binning parameters (Figure 2.11), we noted that the correlation coefficient of specific sample groups decreases as the number of bins is increased. Although this might seem counterintuitive, when considering the signal-to-noise ratio between samples, variability becomes amplified as the number of bins increases.

When compared to the CHUBACA pipeline, clustering the T cell dataset presented in Figure 2.12 using the mean value of each of the 6 parameters the results are subtly, but meaningfully different. In particular, Sample 4 in Figure 2.12, which previously clustered with similarly treated samples, clustered by itself when the mean was used instead of the vectorized flow data (Figure 2.15). We believe this is due to the increased cell-to-cell heterogeneity within that sample; information of this sort is lost when considering only the mean values. Additionally, samples which had returned to baseline levels (bottom three samples in Figure 2.12A) clustered less coherently and was interspersed with other samples but not with the unstimulated controls as expected (cf. blue samples in Figure 2.15A vs. 2.15B). We believe this change is due to the small increase in the mean pSLP76 levels when compared to the unstimulated controls. As described earlier, when mean or median values on whole populations of cells are used to compare samples (Irish et al. 2004), much of the valuable information on individual cells generated by flow cytometry is lost. By considering and utilizing the granularity inherent to the single-cell resolution of flow cytometry data, CHUBACA captures and retains these data for multiple populations and takes cell heterogeneity into account. This is in contrast to mean and median analyses in which the granularity is lost. As with any experiment, it is very important to identify the specific question to be answered when choosing an analysis method to ensure the end result reflects an unbiased analysis of the experimental system and goal.

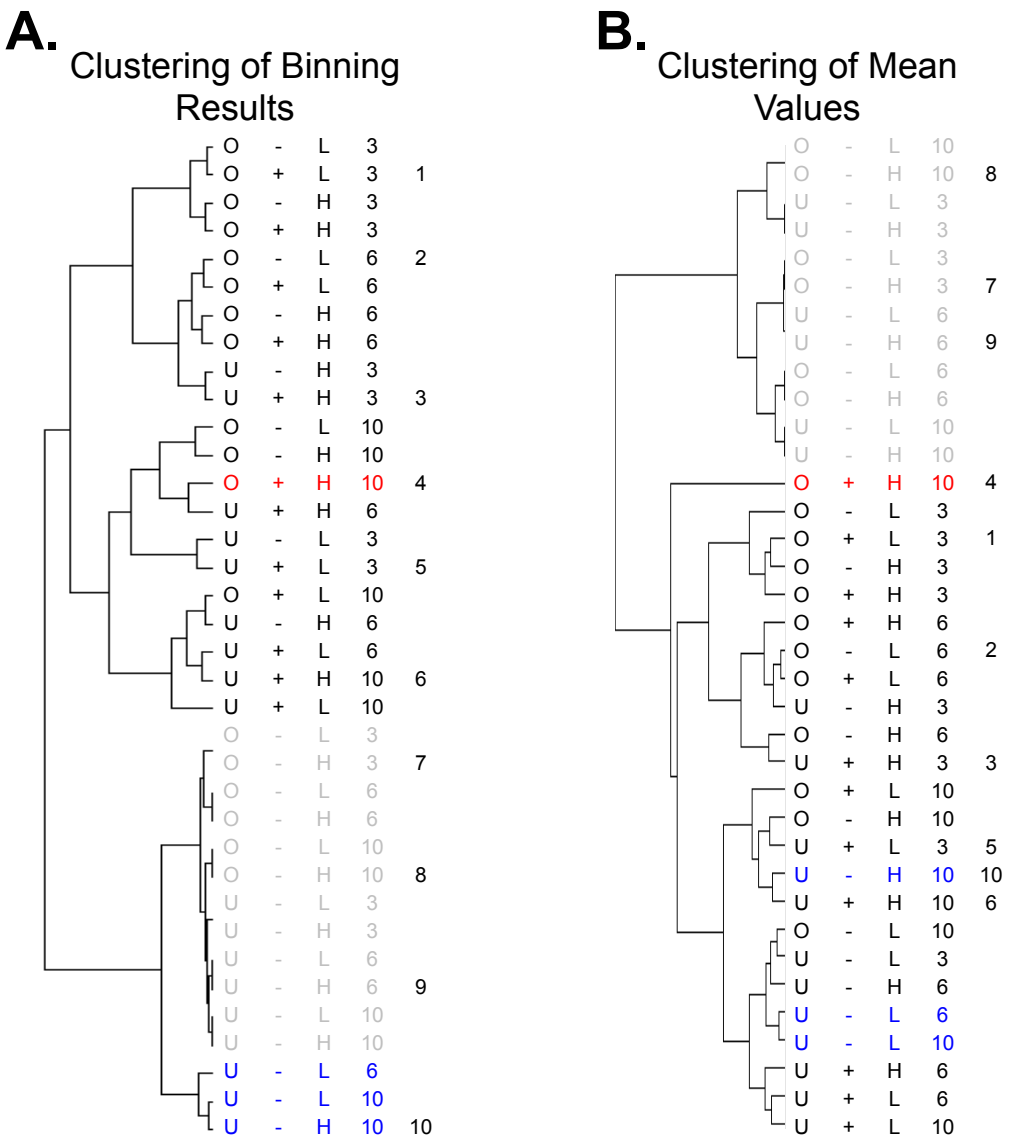


Figure 2.15: Clustering T cell data based on binning versus mean values.

(A) Clustering of binned data from Figure 2.12 (to be used as a reference). The only difference is that additional color-coding was used for specific samples for comparison with panel B (as discussed in main text). See Figure 2.12 legend for a description of the columns. (B) Dendrogram reflecting the relatedness of the 36 T cell samples using the mean values of each of the 6 parameters compared in panel A. Mean values were calculated using FlowJo and represent the same events that were binned in panel A. The red colored sample, Sample 4 (O + H 10), and the blue colored samples have different clustering behavior depending on the methods used for comparison.

One limitation we have encountered is working with an odd number of parameters, since the binning step requires two variables. We have overcome this by simply binning the variable of interest (on the y-axis) with any other parameter (on the x-axis), then collapsing the rows by determining the sum within the respective row (Illustration 2.2). This retains important elements of the single-cell nature of the data and the resulting data can be added to binning results of other parameters, as in Figure 2.12. Another limitation we foresee is comparing data from different cytometers. If systems are calibrated well, direct comparisons can be made if the proper controls cluster with each other from both platforms. If not, a transform can be applied to the scale values to resolve differences between flow cytometers.

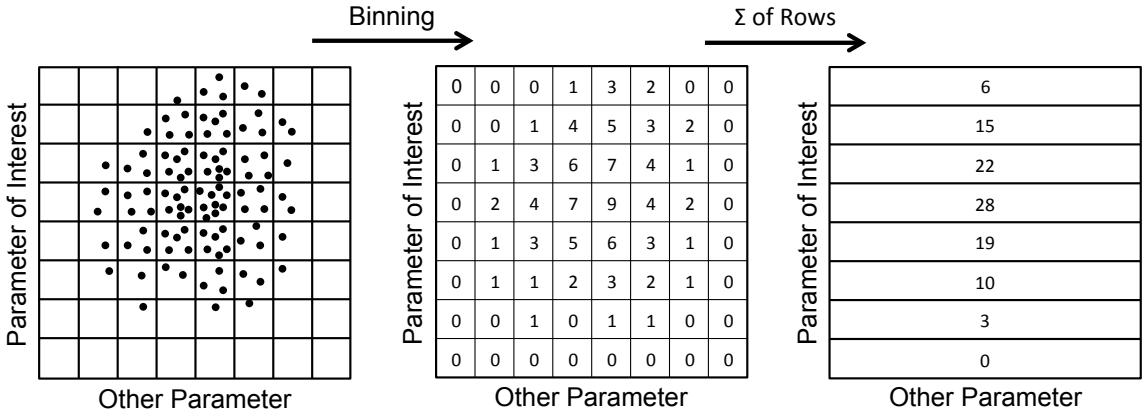


Illustration 2.2: Binning and processing of an additional (odd) parameter

A cartoon two-dimensional plot is shown on the left. A parameter of interest (which will be queried) is plotted on the y-axis and an additional parameter is plotted on the x-axis, in order to employ the binning algorithm. Events are binned as described in the Materials and Methods Section 2.2.5. The middle panel shows the binning results of the panel on the left. To eliminate the "other parameter" and only query the parameter of interest, the total number of events in each row is determined by taking the sum of all bins within a row. The resulting data can then be added to the binning results of the other two-dimensional plots. This process can be automated using a simple script.

Moreover, we can imagine our analysis pipeline being adapted and used to compare other experiments or data that result from multi-variable input, adding to the inherent utility of our system. We trust that CHUBACA will serve as a new analytical tool to explore high content and/or high-throughput flow cytometry data by establishing the relationship of samples based on the similarity of multiple parameters. In contrast to manual analysis of multiple flow cytometry plots, typically performed using 2D-overlays, CHUBACA is semi-automated, decreasing both subject bias and time spent. The visual nature of the analysis and resulting dendrograms are intuitive, quantitative and can be applied to a wide range of experiments. Our analysis pipeline has been used to survey thousands of genetic and chemical perturbations in diverse datasets and has the capability to generate exciting new discoveries with existing and new flow cytometry data.

Although researchers have scoured the yeast genome for decades, it is often with the advent of new technologies and techniques that new breakthroughs occur. Here, I have described the design, validation, and broad utility of our novel gene expression reporter (Sorenson and Stevens 2014). The reporter not only improves upon previous pre-mRNA splicing reporters in terms of speed, sensitivity, and genetic limitations, but it also is sensitive and specific to many other gene expression events. Our analysis pipeline, CHUBACA, coupled with high-throughput flow cytometry allows for rapid, unbiased evaluation of thousands of mutations or conditions. Using our gene expression reporter with standard genetic screening approaches, in series with fluorescence activated cell

sorting allows researchers to explore gene expression processes, and connections within and between them, with increased speed and sensitivity. The broad utility and uniqueness of our reporter make the budding yeast genome, which is often described as barren due to deep cultivation for years, ripe again for identifying novel factors and interactions involved in gene expression.

Chapter 3: Discovery of small molecule modulators of gene expression

3.1 Background

3.1.1 Introduction to drug discovery

Humans have continually been searching for compounds or substances that have the ability to eliminate or alleviate medical symptoms resulting from trauma or an underlying illness (Drews 2000; Sundberg 2000; Pina et al. 2009; Jones 2011). Centuries ago those involved in these searches relied on traditional remedies using natural products that displayed efficacy. Moving forward, early scientific researchers sought to identify the active ingredients of these traditional remedies using crude biochemical techniques. Once an active compound was identified, a common approach is to employ synthetic chemistry to develop derivatives of the compound to optimize the efficacy of the drug. Decades of identifying and generating natural and synthetic small molecules and related derivatives have led to vast chemical databases and libraries, serving as an ideal starting point for most drug discovery efforts. Screening these libraries in a high-throughput manner has been a staple of biomedical and pharmaceutical research (Macarron et al. 2011).

Since the sequencing of the human genome (Lander et al. 2001) and the advent of modern cloning and biochemical techniques, the common approach to drug discovery at the laboratory bench has been reverse pharmacology. This

process involves the high-throughput screening of large libraries of compounds against a purified protein in an *in vitro* reaction to identify inhibitors or modulators of the target protein. It is important to note that compounds that can do so by means other than enzymatic inhibition, such as inhibiting the interaction of a molecule with another, or modulating the localization or stability of that protein. The aforementioned protein-based assays typically utilize colorimetric, fluorometric, chemiluminescent, light scattering, or any other number or combination of methodologies that result in a measurable signal.

Although the standard drug discovery pipeline is a tried and true process, it is very expensive and often takes 10 or more years from initial screens to FDA approval, and can cost as much as $5 billion dollars (DiMasi et al. 2003; Morgan et al. 2011). Large pharmaceutical companies are often abandoning drug screening efforts and either outsourcing the work to academic or research facilities or acquiring promising drug candidates through licensing or acquisitions. These trends, along with heavy competition and intellectual property issues have partially led to a trend of increasing development of orphan drugs (Sharma et al. 2010), which treat rare medical conditions. There are also specific sources of federal and private funding as well that help fuel advancement in this area.

3.1.2 Cell-based drug discovery assays

In comparison with the relatively simple approach using specific purified protein(s) in a biochemical assay, one can also use a cell-based approach to

identify candidate drugs with high-throughput screening. Although this can drastically complicate the screening process in terms of cell growth, contamination, etc., it is most often considered to be advantageous because the assay is completed in the context of a cell, in which a drug would most likely need to be stable, active, and nontoxic. An advantage of cell-based assays is the target is much more likely to be in its native form, conformation, or complex; in contrast to a purified system in which the target might adopt a non-natural shape or conformation. Although cell-based assays have their benefits, as discussed, there is still a high rate of candidate drug failure downstream in the pipeline. The balance of advantages and challenges of such systems in drug discovery are beyond the scope of this work and have been reviewed by others (Astashkina et al. 2012).

Instead of being restrained to the previously mentioned signals of a biochemical assay, researchers can gauge cell viability or any type of measurable cellular response when screening compounds. For example, a common approach in early chemotherapeutic discovery is screening chemical libraries against diseased and normal cells, looking for specific effects or toxicity in the diseased cells. A wide range of human cell lines have been commonly used, with embryonic and induced pluripotent stem cells becoming more routine for certain disease models with increasing feasibility (Sartipy et al. 2007; Crook and Kobayashi 2008; Jensen et al. 2009; Phillips and Crook 2010; Inoue et al. 2014; Engle and Vincent 2014). Cell types can be specifically chosen depending

on the potential target or disease studied. Additionally, panels of different types of human cells are often used to gauge toxicity and specificity in subsequent screening. One specific challenge to cell-based assays is target identification, which is of critical importance since the measured phenotype or response is assayed in the context of biological system (Schenone et al. 2013). There are many methods for identifying drug targets, from standard affinity/biochemical techniques to next generation RNA sequencing (Ong et al. 2009; Rix and Superti-Furga 2009; Ong et al. 2012; Wacker et al. 2012).

3.1.3 Cyclic peptides as therapeutic compounds

A separate class of bioactive small molecules is cyclic peptides. Cyclic peptides are natural-product-like in structure, offering a very attractive chemical space for discovering small molecules that interact with intracellular and extracellular biological targets (Driggers et al. 2008; Thorstholm and Craik 2012; Joo 2012; Bockus et al. 2013). Many natural cyclic peptides are not synthesized ribosomally and rely on large multienzyme complexes that function as an assembly line (Sieber and Marahiel 2003). In terms of biological production of cyclic peptides in the laboratory, there are many different biochemical approaches available (Aboye and Camarero 2012). Examples include using proteases (Marx et al. 2003; Thongyoo et al. 2008), inteins (Tavassoli and Benkovic 2007), thioester-based ligation using modified inteins (Kimura et al.

2006; Austin et al. 2009, 2010) or genetic program recoding (Kawakami et al. 2009).

One specific system used to generate head-to-tail cyclic peptides *in vivo* is SICLOPPS, which stands for Split-Intein-mediated Circular Ligation Of Proteins and PeptideS (Scott et al. 1999; Horswill et al. 2004; Horswill and Benkovic 2005; Kritzer et al. 2009). The genetic construct that generates these peptides uses a split intein approach that allows a variable linker region to be spliced out as a cyclic peptide. The variable linker region can code for a desired stretch of amino acids with a length of anything greater than 4 amino acids. The power of this specialized chemical genetics system first comes from the ability to generate a library of plasmids (and in turn cyclic peptides) by randomizing the linker region, creating libraries that can be screened to identify cyclic peptides of interest (Horswill and Benkovic 2005). Once a clone is identified and verified, structure-activity relationships can be queried by simply employing site-directed mutagenesis to change the structure of the peptide. Moreover, a candidate cyclic peptide can then be synthesized *in vitro* and used in affinity chromatography to identify binding targets.

Although the SICLOPPS system was originally developed in bacteria, it has been successfully adapted to multiple eukaryotic systems. The first scheme involved human B cells in which inhibitors of interleukin-4 signaling were identified from a relatively small library of 270,000 cyclic peptides (Kinsella et al. 2002). SICLOPPS was also adapted for use in budding yeast to screen a library

of 5 million members to rapidly identify two cyclic peptides that specifically reduced the toxicity in a Parkinson's model (Kritzer et al. 2009). Additionally, these two peptides also prevented dopaminergic neuron loss in a *Caenorhabditis elegans* Parkinson's model. A visual description of the SICLOPPS construct and the protein splicing chemistry are shown in Illustration 3.1.

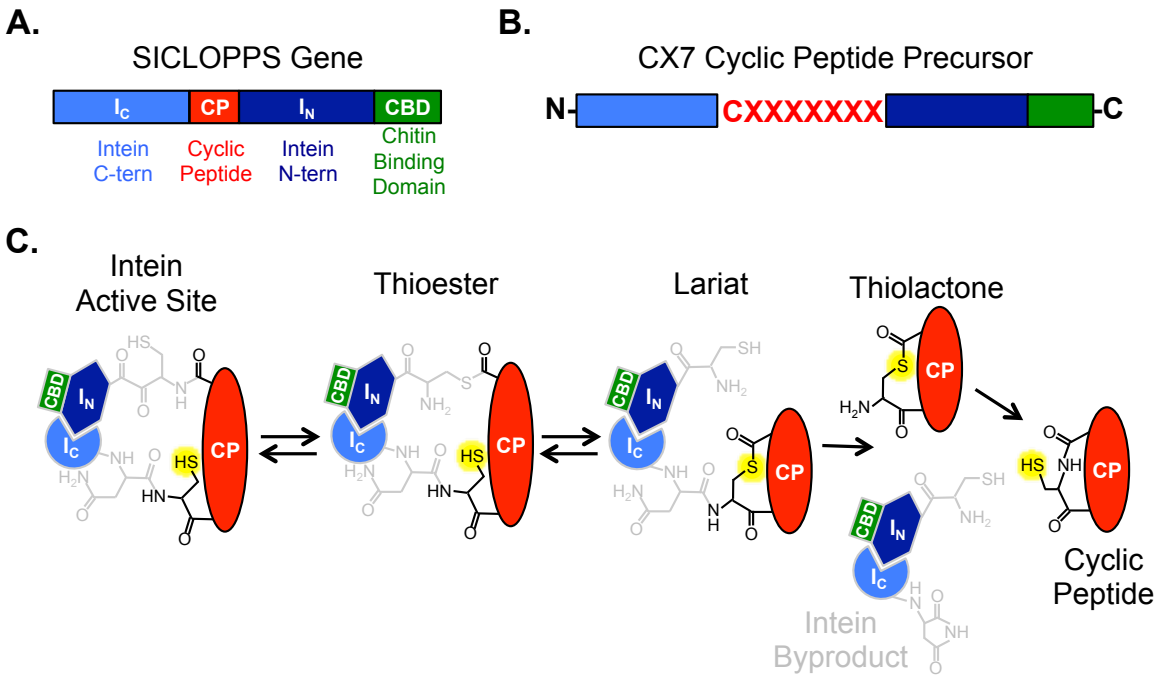


Illustration 3.1: Description of the SICLOPPS construct and splicing mechanism

(A) The SICLOPPS (Split-Intein-mediated Circular Ligation Of Proteins and PeptideS) construct is shown. The *dnaE* self-splicing intein is split into two, and the C-terminal half is placed upstream of the cyclic peptide linker which is followed by the N-terminal protein of *dnaE* and a chitin-binding-domain (CBD) tag for monitoring expression and processing. The strong and constitutive *PGK1* promoter drives expression of the construct. (B) The primary structure of the resulting cyclic peptide precursor is shown, which contains a cysteine followed by 7 random amino acids (X) for the cyclic peptide linker. (C) The circular ligation mechanism chemistry is illustrated. The split-intein first forms a ligase active site and catalyzes an N-to-S acyl shift to produce the thioester. Via a transesterification, the highlighted sulfur acts as a nucleophile, forming the lariat. From there, cyclization with the I_C asparagine side-chain frees the cyclic peptide as a thiolactone and produces the intein byproduct. Finally, an S-to-N acyl shift generates the stable lactam cyclic peptide product. The precursor and intein byproduct can be distinguished by western blot via the CBD. Gray side-chains are part of intein while and black side-chains makeup part of the cyclic peptide. This illustration was adapted from Scott et al. 1999 and aided by personal communication with Joshua Kritzer (Tufts).

Cyclic peptides have been long used as drugs. In 1909 Sir Henry H. Dale discovered and named an unknown substance that contracted the uterus of a pregnant cat. It was not till 1953 that the structure of the hormone oxytocin was discovered (Du Vigneaud et al. 1953), which is used to induce labor. The fungal derived cyclosporine has been used as a immunosuppressant (Kahan 1985). Bacitracin, a mixture of cyclic peptides from *Bacillus subtilis* is used as a antibiotic effective against gram positive and gram negative bacteria (Johnson et al. 1945). Furthermore Actinomycin D (Waksman and Woodruff 1940), which was the first antibiotic that was shown to have anti-cancer activity (Hollstein 1974), inhibits transcription by binding DNA. Although these cyclic peptides can be large and bulky, they are extremely stable in the digestive tract, making them very attractive to designers of protein-based drugs.

3.1.4 Yeast as a tool for drug discovery

The budding yeast *Saccharomyces cerevisiae* has been hailed as a powerful tool for drug discovery (Buschini et al. 2003; Simon and Bedalov 2004; Auerbach et al. 2005; Barberis et al. 2005; Pereira et al. 2012; St Onge et al. 2012; Bilsland et al. 2013; Bancos et al. 2013). There is a high degree of conservation in terms of genes, cellular machinery, and processes between mammalian cells and yeast. Moreover, a large number of human genes have orthologs in lower eukaryotes (Venter et al. 2001). Yeast has many properties that make it a very attractive tool for drug screening. These features include but

are not limited to its short doubling time, inexpensive and simple media requirements, and the ease by which it is genetically modifiable and tractable.

One feature of yeast that can limit its potential in driving the discovery of therapeutics or biological tools is that yeast can be relatively insensitive to compounds that negatively affect other types of eukaryotic cells. The current belief is that this phenomenon is due to the large number of transporters and efflux pumps present in yeast (Paul and Moye-Rowley 2014). The most common approach has been to delete genes that code for transcription factors, such as *PDR1* and *PDR3* (Delaveau et al. 1994; Mamnun et al. 2002), that serve to drive expression of many ABC transporters (Oldham et al. 2008; Locher 2009). Membrane permeability can also be increased by deleting the sterol biogenesis gene *ERG6* (Gaber et al. 1989), rendering yeast sensitive to a wide range of compounds when combined with *pdr1Δ* and *pdr3Δ* (Simon and Bedalov 2004). A curious finding is that Pdr1p and Pdr3p also regulate the expression of hexose transporters, which also can affect pleiotropic drug resistance (Nourani et al. 1997). One group took a brute force approach and deleted 6 ABC transporters in a strain lacking *PDR3*, resulting in a hypersensitive yeast strain referred to as AD12345678 (Decottignies et al. 1998). This idea was taken to another level by Fredrick Roth's lab when they used a creative method of sexual assortment and fluorescence selection termed Green Monster (Suzuki et al. 2011). With this technology they deleted all 16 ABC transporters, generating an even more

sensitive yeast strain (ABC-16) with broad drug-sensitivity, even when compared to AD12345678.

Many approaches have also exploited the expression of human proteins in yeast. Of note, 12 of 38 human cDNAs, which are considered potential chemotherapeutic targets, inhibited growth when expressed in yeast (Tugendreich et al. 2001), providing a means of selection for inhibitors. One interesting tool that was developed is a yeast strain expressing the active form of platelet-derived growth factor receptor β (PDGFRβ), which results in decreased growth rate (Barberis et al. 2005). When these slow growing cells are treated with Gleevec (PDGFRβ inhibitor and chemotherapeutic), growth is restored, creating a situation in which a positive growth selection can be used in screening. Yeast have been successfully used to identify novel human PARP1 inhibitors (Perkins et al. 2001), which were shown to improve the efficacy of DNA-damaging agents used in chemotherapy (Tentori et al. 2003). A multiplexed yeast screen utilizing high-throughput flow cytometry identified a novel inhibitor of yeast TORC1 (Chen et al. 2012). Combinatorial effects of common drugs can also be studied in yeast (Torres et al. 2013). A specific study discovered a suppressive interaction between the antilipemic gemfibrozil and the antidepressant fluoxetine when looking at effects on cell cycle (Hoose et al. 2012).

Yeast has not only been used to identify compounds, but also to identify unknown targets of small molecules, the most famous example being mTOR and

rapamycin (Heitman et al. 1991). Moreover, in taking advantage of the arsenal of genetic tools available for yeast, a field of yeast chemical genomics has emerged. Essentially, the yeast deletion collections and/or overexpression strains are screened in the presence of compounds to identify positive or negative interactions (Giaever et al. 2004; Lum et al. 2004; Parsons et al. 2004; Fortney et al. 2013), helping to deconvolute the mechanism(s) of action for the given compound. Other tools in yeast that have been used for target identification are yeast three-hybrid methods (Griffith et al. 2000; Henthorn et al. 2002), as well as microarray techniques in fission yeast (Arita et al. 2011).

3.1.5 Gene expression machinery as drug targets

Many processes within gene expression are attractive drug targets for diseases such as cancer. Early in gene expression, even before genes are transcribed, nucleosome/chromatin remodeling and histone modification have been shown to be implicated in metastasis and tumorigenesis (Li et al. 2012a; Chernikova et al. 2012). Transcription inhibitors, such as fludarabine and 5-fluorouracil, are used to treat a wide range of cancers (Mandel 1981; Huang et al. 2000). Of particular interest is the increasing number of connections between pre-mRNA and cancer, which has been extensively reviewed (Faustino and Cooper 2003; Srebrow and Kornblihtt 2006; Wang and Cooper 2007; Grosso et al. 2008; Cooper et al. 2009; Oltean and Bates 2013). Of note were separate groups identifying mutations in SF3B1 in myelodysplasia and leukemia (Yoshida

et al. 2011; Papaemmanuil et al. 2011; Rossi et al. 2011; Wang et al. 2011; Malcovati et al. 2011). SF3B1 functions in the catalytic core of the spliceosome, highlighting its importance in gene expression and avoiding a diseased state. More recently, in cases without SF3B1 mutations, PRPF8 (Prp8p in *S. cerevisiae*) mutations conveyed poor patient prognosis and result in a second-step splicing defect (Kurtovic-Kozaric et al. 2014). Moreover, the RNA-binding protein/splicing factor hnRNPA2 is overexpressed in prostate cancer (Stockley et al. 2014) and has also been shown to be involved in liver cancer development (Shilo et al. 2014).

Beyond pre-mRNA splicing, other mRNA processing and transport processes are also involved in avoiding disease. For example, the immunosuppressive drug mizoribine monophosphate blocks the proliferation of B and T lymphocytes, apparently by inhibiting the human RNA capping enzyme (Picard-Jean et al. 2013). The 5' cap structure of mRNAs is important for not only stability and translation of mRNAs, but also their splicing and export. In terms of export, the germinal center-associated protein (GANP), which is an integral part of the mammalian mRNA export pathway (Wickramasinghe et al. 2010), is associated with transformation of melanocytes (Kageshita et al. 2006) and is present in elevated levels in lymphomas (Fujimura et al. 2005). Interestingly, researchers have also found that knockdown of GANP can induce selective death of p53-deficient cancer cells (Phimsen et al. 2012). Additionally, other mRNA export factors, such as CRM1, nucleoporin 88, REF/Aly, and THO are

dysregulated in samples from tumors; which have recently been reviewed (Hurt and Silver 2008; Siddiqui and Borden 2011).

The biogenesis of mRNAs is not the only aspect of gene expression that is important in disease, as posttranscriptional regulation, mRNA stability, and translation are also important to control. One exciting area in cancer research is that of microRNAs, which are most often dysregulated in cancer (Iorio and Croce 2012). Not only are there potential therapeutic targets within this pathway, but also there is also great promise in terms of diagnostics. An essential component to cellular proliferation is translation.

3.1.6 Experimental approach and rationale

We sought to further take advantage of the utility of our gene expression reporter, which is sensitive to defects in many gene expression pathways, by performing complementary screens to identify small molecules that modulate gene expression. Instead of the traditional yeast drug screening approach of looking for growth or absence of growth, our assay is much more diverse and versatile in terms of the potential phenotypes that can be used to identify small molecules of interest. Not only can our approach identify biologically interesting modulators of gene expression, but also it can potentially identify novel targets for future small molecule therapy approaches.

This chapter describes the efforts taken to discover novel small molecules that inhibit or modulate specific gene expression pathways. The first

experimental approach I took was to use the yeast optimized SICLOPPS system in conjunction with our reporter (Kritzer et al. 2009). By using fluorescence-activated cell sorting, I sorted clones that shifted our reporter towards the mCherry or GFP axis separately and identified two clones of interest. Alanine scans were performed to determine structure-activity relationship. Our red-shifted clone, R6_E04, reproducibly shifted reporter expression towards the mCherry axis and mimicked a pre-mRNA splicing defect. RT-qPCR analysis identified minor gene specific pre-mRNA splicing defects of endogenous introns. Affinity chromatography techniques were unsuccessful in determining *in vivo* binding targets, leaving the specific biological activity of this cyclic peptide largely unknown.

The second approach we took was to use our reporter to identify small molecules that modulate gene expression. Our copper-inducible reporter was used in conjunction with a drug-sensitive yeast strain (Suzuki et al. 2011) to perform a pilot screen of a diverse set of small molecules. Out of 3467 initial compounds we tested 354 for dose dependency in a secondary screen and identified 16 compounds of interest. Most of the 16 compounds were green-shifted while only two were significantly red-shifted. I discovered that one of the red-shifted small molecules, TI3D_0104_B19, gives rise to a reproducible and dose dependent pre-mRNA splicing defect. Further experimentation will be needed to determine its mechanism of action. This pilot screen validates our

high-throughput screening approach and lays the foundation for a more broad and extensive screening effort.

3.2 MATERIALS AND METHODS

3.2.1 Cyclic peptide plasmid libraries

The SICLOPPS plasmid libraries were kindly provided by Joshua Kritzer (Kritzer et al. 2009). The four libraries that were obtained were based on high copy *URA3* plasmids (pRS426) with the *PGK1* promoter. The libraries differ in the size of the linker/cyclic peptide; including 5, 6, 7, or 8 random amino acids following a cysteine (named CX5-CX8 and stored as plasmids MS307-310 respectively). A control cyclic peptide, HPQ (MS306), was also provided which is expressed and processed yet has no known phenotypes. The libraries are reported to contain between 30 and 50 million unique clones. Libraries and plasmids discussed can be found in Table 3.1

Table 3.1 Plasmids used in Chapter 3

MS #	Common Name	Purpose	Source
169	pRS415 *TDH3*-Reporter	*TDH3* driven reporter	This work. Plasmids MS 124 and 125
163	pRS316-*CUP1*-Reporter	CEN *URA3* inducible reporter	This work. Plasmids MS 162 and 127.

(Table 3.1 continued)

165	pRS413 *CUP1*-Reporter	CEN *HIS3* inducible reporter	This work. Plasmids MS 101 and 163.
306	pRS426 PGK-HPQ	CFTNVHPQFANA cyclic peptide, used as negative control	Kritzer lab
307	pRS426 PGK-CX5 Library	Cyclic peptide library with 5 random amino acids	Kritzer lab
308	pRS426 PGK-CX6 Library	Cyclic peptide library with 6 random amino acids	Kritzer lab
309	pRS426 PGK-CX7 Library	Cyclic peptide library with 7 random amino acids	Kritzer lab
310	pRS426 PGK-CX8 Library	Cyclic peptide library with 8 random amino acids	Kritzer lab
302	CX7 G5_C07	Green-shifted clone from CX7 screen; CWLCFGCA	Hit from screen, this work
303	CX7 R6_E04	Red-shifted clone from CX7 screen; CSEVGAGL	Hit from screen, this work
304	CX7 G5_C07 ii	Intein inactivated mutant of G5_C07	This work
305	CX7 R6_E04 ii	Intein inactivated mutant of R6_E04	This work
100	pRS413	Yeast expression plasmid	(Sikorski and Hieter 1989)
330	CX7 R6_E04 C1A	Mut. with MS499/500 ASEVGAGL	This work
331	CX7 R6_E04 C1S	Mut. with MS501/502 SSEVGAGL	This work
332	CX7 R6_E04 S2A	Mut. with MS503/504 CAEVGAGL	This work
333	CX7 R6_E04 E3A	Mut. with MS505/506 CSAVGAGL	This work
334	CX7 R6_E04 V4A	Mut. with MS507/508 CSEAGAGL	This work

(Table 3.1 continued)

335	CX7 R6_E04 G5A	Mut. with MS509/510 CSEVAAGL	This work
336	CX7 R6_E04 A6G	Mut. with MS511/512 CSEVGGGL	This work
337	CX7 R6_E04 G7A	Mut. with MS513/514 CSEVGAAL	This work
338	CX7 R6_E04 L8A	Mut. with MS515/516 CSEVGAGA	This work
339	CX7 R6_E04 V4K	Mut. with MS523/524 CSEKGAGL	This work
340	CX7 R6_E04 V4K ctrl 1	Mut. with MS527/528 CSAKGAGL	This work
341	CX7 R6_E04 V4K ctrl 3	Mut. with MS529/530 CSEKGAGA	This work
342	CX7 R6_E04 V4K ctrl 2	Mut. with MS539/540 CSAKGAGA	This work
343	CX7 G5_C07 C1A	Mut. with MS481/482 CWLCFGCA	This work
344	CX7 G5_C07 C1S	Mut. with MS482/483 SWLCFGCA	This work
345	CX7 G5_C07 W2A	Mut. with MS484/485 CALCFGCA	This work
346	CX7 G5_C07 L3A	Mut. with MS486/487 CWACFGCA	This work
347	CX7 G5_C07 C4A	Mut. with MS488/489 CWLAFGCA	This work
348	CX7 G5_C07 F5A	Mut. with MS490/491 CWLCAGCA	This work
349	CX7 G5_C07 G6A	Mut. with MS492/493 CWLCFACA	This work
350	CX7 G5_C07 C7A	Mut. with MS494/495 CWLCFGAA	This work
351	CX7 G5_C07 A8G	Mut. with MS496/497 CWLCFGCG	This work
352	CX7 G5_C07 L3K	Mut. with MS531/532 CWKCFGCG	This work
353	CX7 G5_C07 C7K	Mut. with MS533/534 CWLCFGKG	This work

3.2.2 Development of yeast cyclic peptide library for reporter screen

Standard yeast methods were used (Sherman 1991). For cyclic peptide screening and analysis, yeast strain BY4742 (SS4050) was transformed with pRS415 *TDH3*-Reporter (MS169). This strain was subsequently transformed with the pRS426 *PGK1*-CX7 SICLOPPS library (MS309), which was kindly provided by Joshua Kritzer at Tufts University (Kritzer et al. 2009), and plated on synthetic dropout medium lacking leucine and uracil. The construct pRS426 *PGK1*-HPQ (MS317) was also provided and transformed into the same strain to use as a negative control and for sorting purposes. The resulting transformants were pooled by washing cells off of plates with water. Cells that were not immediately analyzed were cryopreserved in 25% glycerol at -80°C. Strains used in Chapter 3 can be found in Table 3.2

Table 3.2 Yeast strains used in Chapter 3

Strain (SS)	Mating type	Genotype
4050	*α*	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* (BY4742)
4127	*α*	*(adp1Δ, aus1Δ, bpt1Δ, nft1Δ, pdr5Δ, pdr10Δ, pdr11Δ, pdr12Δ, pdr15Δ, snq2Δ, vmr1Δ, ybt1Δ, ycf1Δ, ynr070wΔ, yol075cΔ, yor1Δ); at each loci, tetO2-GFP::URA3 to replace KanMX) can1Δ::GMToolkit-alpha (CMVpr-rtTA NatMX4 STE3pr-LEU2) lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*
4136	*α*	*(adp1Δ, aus1Δ, bpt1Δ, nft1Δ, pdr5Δ, pdr10Δ, pdr11Δ, pdr12Δ, pdr15Δ, snq2Δ, vmr1Δ, ybt1Δ, ycf1Δ, ynr070wΔ, yol075cΔ, yor1Δ); at each loci, tetO2-GFP::URA3 to replace KanMX) can1Δ::GMToolkit-alpha (CMVpr-rtTA NatMX4 STE3pr-LEU2) lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0* + pRS413 *CUP1*-Reporter

3.2.3 Fluorescence-activated cell sorting for cyclic peptide screen

To prepare the yeast culture that would generate cells to sort, the pooled library was used to inoculate 10 mL of synthetic dropout medium lacking leucine and uracil at OD (A_{600}) 0.1. The culture was incubated at 31°C for 7-8 hours. Immediately prior to flow cytometry, the cells were harvested by centrifugation and resuspended in sterile 1X PBS. For fluorescence-activated cell sorting (FACS), the HPQ control (plasmid MS306) was used to determine gating boundaries. Two gates were created on the prototypical mCherry versus GFP plot (y-axis and x-axis respectively), one to the left of the control population (red-shifted) and one to the right (green-shifted). Gates were placed in order to sort between 0.1 and 0.3% of the total population that was shifted most significantly towards the respective axis. 10,000+ cells were sorted from each gate and half the cells were plated for single colonies on large format synthetic dropout agar plates lacking leucine and uracil, while the other half was cryogenically preserved. Individual clones were transferred to flat bottom 96-well plates along with the HPQ controls, cultured, and cryogenically preserved.

3.2.4 Yeast plasmid mini prep

Plasmids from cyclic peptide clones of interest were isolated using the Zymoprep Yeast Plasmid MiniprepII kit (Zymo). Briefly, clones were grown in synthetic dropout media lacking uracil to specifically select for the cyclic peptide plasmid, not the reporter. The protocol for liquid cultures was followed. The

resulting plasmid DNA was transformed into *E. coli* and miniprepped. After a diagnostic digest to make sure the plasmid looked correct, it was sequence verified with oligonucleotide MS480 (Table 3.3).

3.2.5 Western blot analysis

Protein samples were separated on PAGE gels (Invitrogen) and transferred to nitrocellulose. Membranes were blocked with BLOTTO (PBST + 5% SDS) and then probed with anti-CBP at 1:5000 (NEB, E8034S). Blots were then probed with HRP conjugated goat anti-mouse (Rockland, 610-1302). Washes were preformed with PBST and visualized using X-ray film using chemiluminescence (Perkin Elmer).

3.2.6 Serial dilutions and spot plating

Yeast were grown in the appropriate media to log phase. All cultures were normalized to OD (A_{600}) 1.0 in sterile ddH_2O. In a 96-well plate, five 10-fold dilutions were made. 5 µL of the undiluted and 5 dilutions were spotted onto appropriate media.

3.2.7 Site-directed PCR mutagenesis

For full method refer to Section 2.2.2. Primers for the intein inactivation mutant T69A H72A (Ghosh et al. 2001; Kritzer et al. 2009) were MS478 and MS479. DNA oligonucleotides used in this Chapter can be found in Table 3.3

Table 3.3 DNA oligonucleotides used in Chapter 3

MS#	Sequence (5' to 3')	Purpose
282	CCATTGTCACTCTAGATGTCAAGCC	*TEF5* total FWD qPCR
283	GATACCGAAACCAATTGGGATAAAT	*TEF5* total REV qPCR
295	TCCAGGAGCGCACCATCTT	Reporter total FWD qPCR
296	TCTGCTTGTCGGCCATGATA	Reporter total REV qPCR
299	NNNNNNNNG	NonamerG for RT
300	NNNNNNNNA	NonamerA for RT
301	NNNNNNNNT	NonamerT for RT
302	NNNNNNNNC	NonamerC for RT
307	ACACCATCGTTGAACAATACGAA	Reporter pre-mRNA FWD qPCR
308	TACGACCTTAACCGGCTGTACA	Reporter pre-mRNA REV qPCR
313	GATAGCACAGAGCAGAGTATCATTA	*TEF5*intFWDjp qPCR
314	CTGGAGAATTCTGGGTAAGCAGATT	*TEF5*intREVjp qPCR
478	GCTCAGTAATCCGAGCTGCCTCTGACGCCCGCTTTTTAACC	T69A H72A intein inactivation mutant TOP
479	GGTTAAAAAGCGGGCGTCAGAGGCAGCTCGGATTACTGAGC	T69A H72A intein inactivation mutant BTM
480	GCTGCCACAAGGCAGGAACGTTGGA	Sequencing cyclic peptide clones
481	GGGCGATCGCCCACAATGCTTGGCTGTGCTTC	CX7_G5_C07 Alanine Scan; C0A TOP
482	GAAGCACAGCCAAGCATTGTGGGCGATCGCCC	CX7_G5_C07 Alanine Scan; C0A BTM
483	GGGGCGATCGCCCACAATTCATGGCTGTGCTT	CX7_G5_C07 Alanine Scan; C0S TOP
484	AAGCACAGCCATGAATTGTGGGCGATCGCCCC	CX7_G5_C07 Alanine Scan; C0S BTM
485	CGATCGCCCACAATTGTGCTCTGTGCTTCGGGTGCG	CX7_G5_C07 Alanine Scan; W1A TOP
486	CGCACCCGAAGCACAGAGCACAATTGTGGGCGATCG	CX7_G5_C07 Alanine Scan; W1A BTM
487	CGCCCACAATTGTTGGGCATGCTTCGGGTGCGCGT	CX7_G5_C07 Alanine Scan; L2A TOP
488	ACGCGCACCCGAAGCATGCCCAACAATTGTGGGCG	CX7_G5_C07 Alanine Scan; L2A BTM
489	CCCACAATTGTTGGCTGGCATTCGGGTGCGCGTGCTT	CX7_G5_C07 Alanine Scan; C3A TOP

(Table 3.1 continued)

490	AAGCACGCGCACCCGAATGCCAGC CAACAATTGTGGG	CX7_G5_C07 Alanine Scan; C3A BTM
491	CCACAATTGTTGGCTGTGCGCAGG GTGCGCGTGCTTAAGTT	CX7_G5_C07 Alanine Scan; F4A TOP
492	AACTTAAGCACGCGCACCCTGCGC ACAGCCAACAATTGTGG	CX7_G5_C07 Alanine Scan; F4A BTM
493	AATTGTTGGCTGTGCTTCGCATGCG CGTGCTTAAGTTTTG	CX7_G5_C07 Alanine Scan; G5A TOP
494	CAAAACTTAAGCACGCGCATGCGAA GCACAGCCAACAATT	CX7_G5_C07 Alanine Scan; G5A BTM
495	GTTGGCTGTGCTTCGGGGCCGCGT GCTTAAGTTTTG	CX7_G5_C07 Alanine Scan; C6A TOP
496	CAAAACTTAAGCACGCGGCCCCGA AGCACAGCCAAC	CX7_G5_C07 Alanine Scan; C6A BTM
497	GCTGTGCTTCGGGTGCGGTTGCTT AAGTTTTGGCAC	CX7_G5_C07 Alanine Scan; A7G TOP
498	GTGCCAAAACTTAAGCAACCGCACC CGAAGCACAGC	CX7_G5_C07 Alanine Scan; A7G BTM
499	GGCGACCGCCCACAATGCTTCGGA GGTCGG	CX7_R6_E04 Alanine Scan; C0A TOP
500	CCGACCTCCGAAGCATTGTGGGCG GTCGCC	CX7_R6_E04 Alanine Scan; C0A BTM
501	GGGCGACCGCCCACAATTCTTCGG AGGTC	CX7_R6_E04 Alanine Scan; C0S TOP
502	GACCTCCGAAGAATTGTGGGCGGT CGCCC	CX7_R6_E04 Alanine Scan; C0S BTM
503	CGACCGCCCACAATTGTGCTGAGG TCGGCGCG	CX7_R6_E04 Alanine Scan; S1A TOP
504	CGCGCCGACCTCAGCACAATTGTG GGCGGTCG	CX7_R6_E04 Alanine Scan; S1A BTM
505	CCACAATTGTTCGGCCGTCGGCGC GGGCC	CX7_R6_E04 Alanine Scan; E2A TOP
506	GGCCCGCGCCGACGGCCGAACAAT TGTGG	CX7_R6_E04 Alanine Scan; E2A BTM
507	TTGTTCGGAGGCCGGCGCGGGCC	CX7_R6_E04 Alanine Scan; V3A TOP
508	GGCCCGCGCCGGCCTCCGAACAA	CX7_R6_E04 Alanine Scan; V3A BTM
509	TTCGGAGGTCGCCGCGGGCCTGT	CX7_R6_E04 Alanine Scan; G4A TOP

(Table 3.1 continued)

510	ACAGGCCCGCGGCGACCTCCGAA	CX7_R6_E04 Alanine Scan; G4A BTM
511	TTCGGAGGTCGGCGGAGGCCTGTG CTTAAG	CX7_R6_E04 Alanine Scan; A5G TOP
512	CTTAAGCACAGGCCTCCGCCGACC TCCGAA	CX7_R6_E04 Alanine Scan; A5G BTM
513	CGGAGGTCGGCGCGGCTCTGTGCT TAAGTTTTG	CX7_R6_E04 Alanine Scan; G6A TOP
514	CAAAACTTAAGCACAGAGCCGCGC CGACCTCCG	CX7_R6_E04 Alanine Scan; G6A BTM
515	GGAGGTCGGCGCGGGCGCATGCTT AAGTTTTGGCA	CX7_R6_E04 Alanine Scan; L7A TOP
516	TGCCAAAACTTAAGCATGCGCCCG CGCCGACCTCC	CX7_R6_E04 Alanine Scan; L7A BTM
523	CCACAATTGTTCGGAGAAAGGCGC GGGCCTGTGCT	CX7_R6_E04 V4K mutation
524	AGCACAGGCCCGCGCCTTTCTCCG AACAATTGTGG	CX7_R6_E04 V4K mutation
527	GCCCACAATTGTTCGGCCAAAGGC GCGGGCCT	R6_E04 pull-down control mutation
528	AGGCCCGCGCCTTTGGCCGAACAA TTGTGGGC	R6_E04 pull-down control mutation
531	GCCCACAATTGTTGGAAGTGCTTC GGGTGCGC	G5_C07 L3K mutation
532	GCGCACCCGAAGCACTTCCAACAA TTGTGGGC	G5_C07 L3K mutation
533	ATTGTTGGCTGTGCTTCGGGAAAGC GTGCTTAAGTTTTGGCA C	G5_C07 C7K mutation
534	GTGCCAAAACTTAAGCACGCTTTCC CGAAGCACAGCCAACAAT	G5_C07 C7K mutation
539	GCCCACAATTGTTCGGCCAAAGGC GCGGGCGC	R6_E04 pull-down control mutation
540	GCGCCCGCGCCTTTGGCCGAACAA TTGTGGGC	R6_E04 pull-down control mutation
541	AATCGTGTGACAACAACAGCC	*PGK1* sequencing primer
543	AGAGGTCACGTCTCAGG	*CYH2* Northern probe
544	ACAACACCACCAGCAGC	*CYH2* Northern probe
584	ACAGAATAACGAACCAAATTACTAA CAGT	*DBP2*intFWDms qPCR
585	CCTGGCATATCGTAGTTGATAACG	*DBP2*intREVms qPCR

(Table 3.1 continued)

586	CTTCACCGAACAAAACAAAGGTT	*DBP2*exonFWDms qPCR
587	TCGGGAGGAATATTTTGATTAGCT	*DBP2*exonREVms qPCR
592	TTACAATATTTTTTTTGTACAGCCGG	*CYH2* pre-mRNA FWD qPCR
593	CACCGGCCATACCTCTACCA	*CYH2* pre-mRNA REV qPCR
594	TTATTTCGGTAAGGTTGGTATGAGA TAC	*CYH2* total FWD qPCR
595	TGGGATCAATGTCCACAATTTG	*CYH2* total REV qPCR
598	CCAGACATTTTTTCCTCGTCCAAAC G	*RPL22B* pre-mRNA FWD
599	ACTTTTTGCTTTCTGGAAGTCTGTTA CAG	*RPL22B* pre-mRNA REV
600	TGTAGTATCCAGTGCCAAATTTTCT GG	*RPL22B* total FWD qPCR
601	ACGAATCTGATCCAGTCTCTTAGCT G	*RPL22B* total REV qPCR
602	CGCACCGTGCCCTGTT	*SCR1* FWD qPCR
603	AGCTCTGCCCAGGACAAATTT	*SCR1* REV qPCR
604	GGTAGAACCGCTTCCGGTAAC	*TDH3* FWD qPCR
605	AAGACCTTACCGACAGCCTTAGC	*TDH3* REV qPCR
620	AGGCCGAATTAACTCTACGC	*RPL18B* pre-mRNA FWD qPCR
621	CATGGAGGGACGATAAATATGAA	*RPL18B* pre-mRNA REV qPCR

3.2.8 Pull-down with biotinylated cyclic peptide

Wild-type BY4742 (SS4050) yeast cells were cultured and harvested at OD 1.0 (A_{600}). Cells were washed with cold water and resuspended in 1 mL/gram AGK (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 8% glycerol). Yeast were frozen in liquid nitrogen and subsequently lysed cryogenically in a Retch ball mill (Mayas et al. 2006). The grindate was thawed and centrifuged in a JA-14 rotor (12,000 rpm, 30 minutes, 4°C). Supernatant was transferred to

ultracentrifuge tubes. Extract was centrifuged in a Ti45 (42,000 rpm, 75 minutes, 4°C). The supernatant was dialyzed twice in 4 L of Buffer D20 (20% glycerol, 20 mM HEPES, pH 7.9, 50 mM KCl, 0.5 mM DTT, 1.5 mM $MgCl_2$) for 2 hours at 4°C. Dialyzed extract was transferred to Oakridge tubes and centrifuged in a JA17 rotor (14,000 rpm, 10 minutes, 4°C). 10 mL supernatant was aliquoted to 50 mL falcon tubes, flash frozen in liquid nitrogen, and stored at -80°C.

Extract was thawed in ice water. While thawing, 100 µL NeutrAvidin Resin streptavidin slurry (Thermo) was prepared in a disposable column by washing with IPP150 (10 mM Tris, pH 8.0, 150 mM NaCl, 0.1% NP-40, 1.5 mM $MgCl_2$, 8% glycerol, 1 mM DTT). When extract was thawed, 1 volume of IPP150 was added. Extracts were adjusted to 0.2% NP-40, 1 µM DTT, 0.4 mM PMSF, 5 µM leupeptin, and 3 µM pepstatin. Prepared streptavidin beads were transferred to the extract and rotated for 1 hour at 4°C. Extract and beads were filtered through the disposable column to clear extract. The extract was then cleared again with another 100 µL streptavidin slurry. The twice-cleared extract was collected in a fresh tube, 100 µg cyclic peptide (or none for control) was added, then rotated for 60 minutes at 4°C. Freshly washed streptavidin beads (50 µL slurry) were transferred to the extract with cyclic peptide and the mixture was rotated for 1 hour. Extract was filtered through a disposable column and washed four times with 10 mL IPP150. The washed beads resuspended with 500 µL 0.1 M glycine HCl pH 2.8 and filtered into a microfuge tube containing 50 µL 1M Tris pH 9.0. Proteins were extracted with 1 volume acid-phenol chloroform. The organic

phase was precipitated with 800 µL cold acetone, overnight at -20°C. Pellets were washed twice with cold acetone, twice with 80% ethanol, and allowed to dry at room temperature. Samples were resuspended in 15 µL 2X LDS buffer containing 200 mM DTT and electrophoresed through an SDS-PAGE gel. The gel was fixed, stained with coomassie blue, destained, and visualized on a light box. The other half of the protein samples was saved for MS/MS analysis by the ICMB proteomics core facility.

3.2.9 Generation of random-primed DNA probes

The protocol here was adapted from the instructions provided (Life Technologies). A DECAPrime II Kit was used for a 25 µL reaction. In a microfuge tube, 25 ng of DNA template was added to a total volume of 11.5 µL (adjusted with ddH_2O). For template, gel purified PCR product was used, generated from wild-type yeast genomic DNA and primers MS543 and MS544 (for probing *CYH2*; sequences in Table 3.3). 2.5 µL 10X Decamer Solution was added to the template, mixed, and heated at 95-100°C for 5 minutes. The mixture was flash frozen in liquid nitrogen and placed on ice to thaw (centrifuge to collect). On ice, added 5 µL 5X Reaction Buffer (-dATP), 2.5 µL ddH_2O, 2.5 µL [α-^{32}P] dATP (20 mCi/mL), 1 µL Exo-Klenow. Tube was flicked to mix, spun down, and incubated at 37°C for 10 minutes. Reaction was stopped by the addition of 1 µL 0.5 M EDTA. 25 µL of TE was added. Unincorporated radionucleotides were removed

by running the solution through a sephadex column. Probes were stored for up to one week at -20°C

3.2.10 Northern blot analysis

RNA (15 µg) was dried down using a Speedvac. Dried RNA was then denatured and electrophoresed on a 1% agarose, 2.2 M formaldehyde gel (Sambrook et al. 1989). rRNA was visualized by ethidium bromide staining. RNA was then transferred to a nylon membrane (Whatman) via downward capillary transfer (using 0.01 N NaOH and 3 M NaCl). After crosslinking by UV light (254 nm for 1 minute), the membrane was probed with the random-primed body labeled DNA (boiled immediately prior to addition) in Church Buffer (500 mM Na_2HPO_4 pH 7.0, 7% SDS, 1 mM ETDA pH 8.0) overnight. The blot was then exposed to a phosphorimaging screen and developed in a phosphorimager.

3.2.11 Development of drug screening strain

To develop a compatible reporter for the drug-sensitive strain, the reporter sequence from plasmid pRS316 *CUP1*-Reporter (*URA3* marked, MS163) was removed with BamHI and SalI and ligated into pRS413 (MS100) to generate pRS413 *CUP1*-Reporter (MS165), a low copy *HIS3* marked copper inducible reporter. This reporter was transformed using standard lithium acetate method (Gietz and Woods 2002) into the drug-sensitive green monster ABC-16 strain (ABC-16; SS4127), which was kindly provided by the Roth lab (Suzuki et al.

2011). This strain (SS4136) was cultured in low fluorescent synthetic dropout media lacking histidine and used for drug screening efforts.

3.2.12 Preparation of low-fluorescent synthetic dropout media

The protocol for making this media was adapted from previously published work (Sheff and Thorn 2004). A 200X master stock is prepared with the following amounts of solid chemicals for 500 mL; 50 mg boric acid, 4 mg $CuSO_4$, 10 mg KCl, 20 mg $FeCl_3$, 40 mg $MnSO_4$, 20 mg Na_2MoO_4, 40 mg $ZnSO_4$, 200 mg inositol, 40 mg niacin, 20 mg 4-aminobenzoic acid (PABA), 40 mg pyridoxine HCl, 40 mg thiamine, 40 mg calcium pantothenate. These solids were dissolved in 500 mL ddH_2O, filter sterilized, and stored at 4°C covered with tinfoil. To prepare 500 mL of liquid media, 100 µL 5000X biotin (10 mg/L in ddH_2O, sterile filtered, -20°C) and 2.5 mL of 200X master stock were added to 400 mL dH_2O. While stirring, 2.5 g NH_4SO_4, 0.5 g KH_2PO_4, 0.25 g $MgSO_4$, 50 mg NaCl, and 50 mg Ca_2Cl were added, as was the desired amino acid mixture. Commercial nitrogen base and ammonium sulfate are not added. The solution is then brought up to 500 mL with dH_2O and autoclaved.

3.2.13 Calculation of z-score for plate-based reporter assay

In the validation process of our reporter assay, the Z-factor/Z-prime (Z') was calculated to determine the effect size between our negative control (copper and DMSO to induce reporter to 100%) and our positive control (no copper and

DMSO to inhibit expression of the reporter completely, i.e. 100% inhibition). The calculation used is:

$$Z' = 1 - [3(\sigma_p + \sigma_n) / (|\mu_p - \mu_n|)]$$

σ_p = standard deviation of positive control

σ_n = standard deviation of negative control

μ_p = mean of positive control

μ_n = mean of negative control

The ideal Z' is 1. An excellent assay has a Z' between 0.5 and 1.0 and marginal assays have values between 0 and 0.5. Assays with a Z' of less than 0 are not useful due to not enough separation between positive and negative control or very high levels of deviation between replicates.

3.2.14 High-throughput drug screening

The drug-sensitive yeast strain ABC-16 (Suzuki et al. 2011) with pRS413 *CUP1*-Reporter (MS165; strain SS4126) was grown overnight in synthetic dropout media lacking histidine. This culture was then used to inoculate four 200 mL cultures with the same media at OD (A_{600}) 0.0005, 0.00075, 0.001, and 0.0015 in 1 L-baffled culture flasks. These cultures were grown for 16 hours at 31°C in order to have one culture as close to OD (A_{600}) 1.0 as possible (OD (A_{600}) 0.9 to 1.1 is ideal). Immediately before bringing the cells and other materials to Ti-3D (Texas Institute for Drug & Diagnostic Development, The University of Texas at Austin) for high-throughput screening, the appropriate

volume of culture was centrifuged in a 50 mL conical in order to have 50 mL of cells at OD (A_{600}) 0.25 and the media was decanted (multiple aliquots of cells were typically needed, 25 mL final volume of culture was needed for each 96-well plate). 50 mL of low fluorescent dropout media lacking histidine was added to the tubes. Cells and other materials were transferred to Ti-3D. Using robotics, 135 µL of cell culture was added to each well of a black 96-well plate with a flat clear bottom. For screens, 1 or 5 µL of stock compounds in DMSO was added to each well, resulting in a final concentration of 10 µM. Cells with compounds were incubated at 31°C on a microplate shaker for 30 minutes to allow compounds to elicit at affect. After 30 minutes, 10 µL 7.5 mM $CuSO_4$ or 15 µL 5 mM $CuSO_4$ (depending on volume of compound) was added to induce reporter expression with a final concentration of 0.5 mM $CuSO_4$, resulting in a total volume of 150 µL/well. Plates were incubated at 31°C for 6 hours. Absorbance, as well as GFP and mCherry fluorescence was measured with a Biotek Synergy H4 plate reader. Secondary screens were preformed as described but with final compound concentrations of 10 µM, 1 µM, and 100 nM.

The first and last column of each plate was not treated with compounds, and one was not treated with copper, to serve as controls. The wells without copper and compound served as a “100% inhibition” positive control to set the baseline fluorescence value. The wells with copper, but without compound were considered to be the “negative control” in which maximum fluorescence was expected. Fluorescence values for each well were divided by the absorbance

readings to normalize to cell density. These normalized fluorescence values were compared to the average of the negative controls to determine % inhibition for each fluorescence value separately. Alternatively, the mCherry/GFP ratio can be used as well. A graphical representation of the screening process is shown in Illustration 3.2.

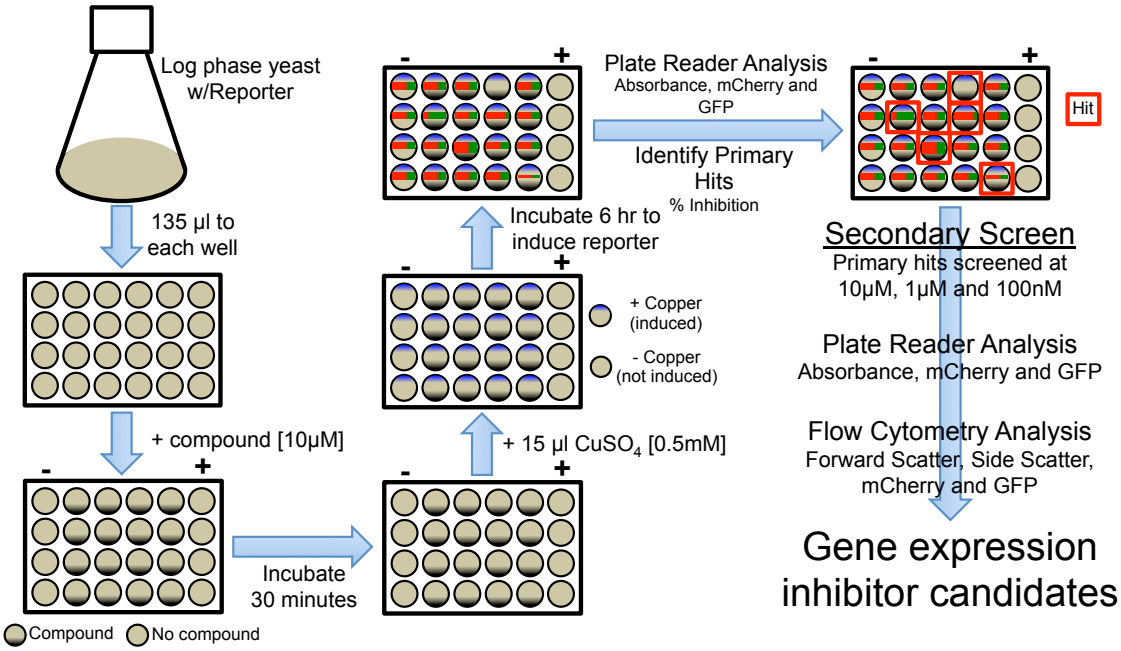


Illustration 3.2: Graphical representation of high-throughput screen pipeline

The general process of our high-throughput small molecule screening efforts is shown. Starting in the top left, drug-sensitive yeast harboring a copper-inducible reporter were dispensed robotically to 96-well plates and compounds were added to a final concentration of 10 µM. After 30 minutes incubation, $CuSO_4$ was added to 0.5 mM and reporter was induced for 6 hours. 96-well plates were rapidly analyzed by a plate reader to identify primary hits, which were then analyzed in a secondary screen in which compounds were tested again at 10 µM, as well as 1 µM and 100 nM. Plate reader and flow cytometry analysis was used to identify candidate gene expression inhibitors or modulators. More specific details can be found in Section 3.2.14.

3.3 RESULTS

3.3.1 FACS-guided cyclic peptide screen results in two high-confidence hits

The first cyclic peptide screen that I performed utilized the CX7 library (MS307) that generates cyclic peptides with 8 amino acids, a cysteine followed by 7 random amino acids. This library chosen to use first based on high quality results from our collaborators with this particular library (the Kritzer Laboratory, Tufts). By performing multiple standard lithium acetate transformations (Gietz and Woods 2002) with 500 ng library plasmid DNA into a wild-type yeast strain harboring our constitutive gene expression reporter, I obtained an estimated 2 million transformants. Over 15 million cells were then analyzed by flow cytometry to sort red- and green-shifted cells. To determine the gates by which the cells were sorted, a control strain was using that harbored both the gene expression reporter and the HPQ control cyclic peptide construct. According to our collaborator, this cyclic peptide expresses well and had no identifiable phenotypes. The reporter phenograph for HPQ expressing cells and the gating boundaries can be seen in Figure 3.1A.

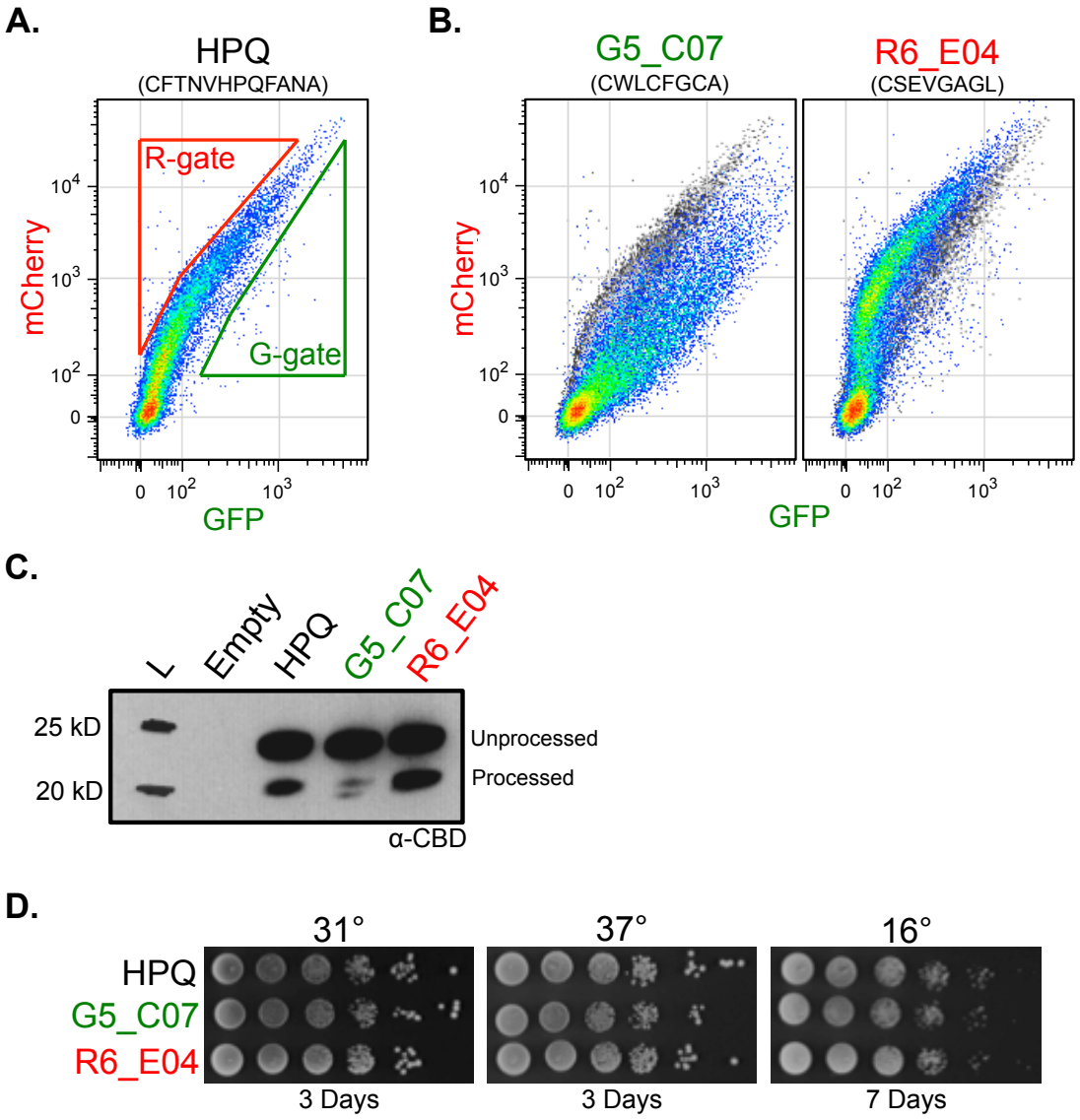


Figure 3.1: Initial characterization of two CX7 cyclic peptide hits

(A) Reporter expression in wild-type yeast with the HPQ control cyclic peptide. The red (R) and green (G) gates used to sort the CX7 library are shown. (B) Reporter phenograph overlays of the two cyclic peptide hits (pseudo color) are shown in comparison to wild-type (grayscale). (C) Western blot analysis, probing for the chitin-binding-domain fused to the SICLOPPS intein, is shown for the HPQ control and the two hits. The slower migrating band is the cyclic peptide precursor and the faster band is the intein byproduct (Illustration 3.1). (D) Serial dilutions of wild-type yeast harboring the indicated cyclic peptide plasmid were made, plated onto synthetic dropout medium, and incubated at the indicated temperature for 3 or 7 days.

Utilizing FACS I collected 91,250 red-shifted cells and 50,096 green-shifted cells. While most of these cells were stored for future analysis, a batch of nearly 2000 were plated for single colonies and analyzed by high-throughput flow cytometry in 96-well plates to identify clones with reproducible reporter phenotypes. By using CHUBACA, our binning and clustering algorithm described in Chapter 1, I identified 66 clones that clustered separately from the HPQ controls. Since we have observed 96-well analysis artifacts in the past, I grew these clones in 5 mL culture tubes and analyzed them in early log phase. Within these clones, there were 18 that resulted in reproducible reporter phenotypes. The plasmids from these strains were isolated from yeast, transformed into *E. coli*, then retransformed into the wild-type yeast strain containing our reporter. Out of the 18 clones that were analyzed, only 2 resulted in a reproducible reporter phenotype, which are shown in Figure 3.1B.

The green-shifted hit, G5_C07 (CWLCFGCA), results in a reporter phenograph that is drastically shifted towards the GFP axis as a result of a decrease in both florescence levels (Figure 3.1B). This phenotype is unlike any other I had seen with the green-shifted deletion mutants (Figure 2.4). The red-shifted hit, R6_E04 (CSEVGAGL), looked very similar to the reporter phenograph from cells with a pre-mRNA splicing defect or those lacking functional NMD (Figure 2.3B and 2.3C respectively). To verify that these clones were indeed expressing the SICLOPPS construct, I performed a western blot to detect the levels of cyclic peptide precursor and processed intein, by virtue of the chitin-

binding-domain tag (Illustration 3.1). As predicted, both G5_C07 and R6_E04 were expressed and processed (Figure 3.1C), albeit G5_C07 appeared to be processed less efficiently than R6_E04 and the HPQ control. Although these clones had obvious reporter phenotypes, they did not have a significant effect on growth rate at permissive, high, and low temperatures (Figure 3.1D).

To determine if processing of the cyclic peptide precursor was required for the reporter phenotype, I introduced the T69A H72A intein inactivation mutation (ii) into these clones to abolish the generation of cyclic peptide (Ghosh et al. 2001). This mutation permits expression of the cyclic peptide precursor but interferes with the active site, resulting in 100% unprocessed cyclic peptide precursor. Western blot analysis verified that the ii mutations completely blocked the processing of both cyclic peptides (Figure 3.2A). As anticipated, the ii mutation also eliminates the reporter phenotypes for both hits (Figure 3.2B). With these data in hand, we were confident that the cyclic peptide elicited the red- and green shifts for clones R6_E04 and G5_C07 respectively. The primary structure of both cyclic peptides is shown in Illustration 3.3.

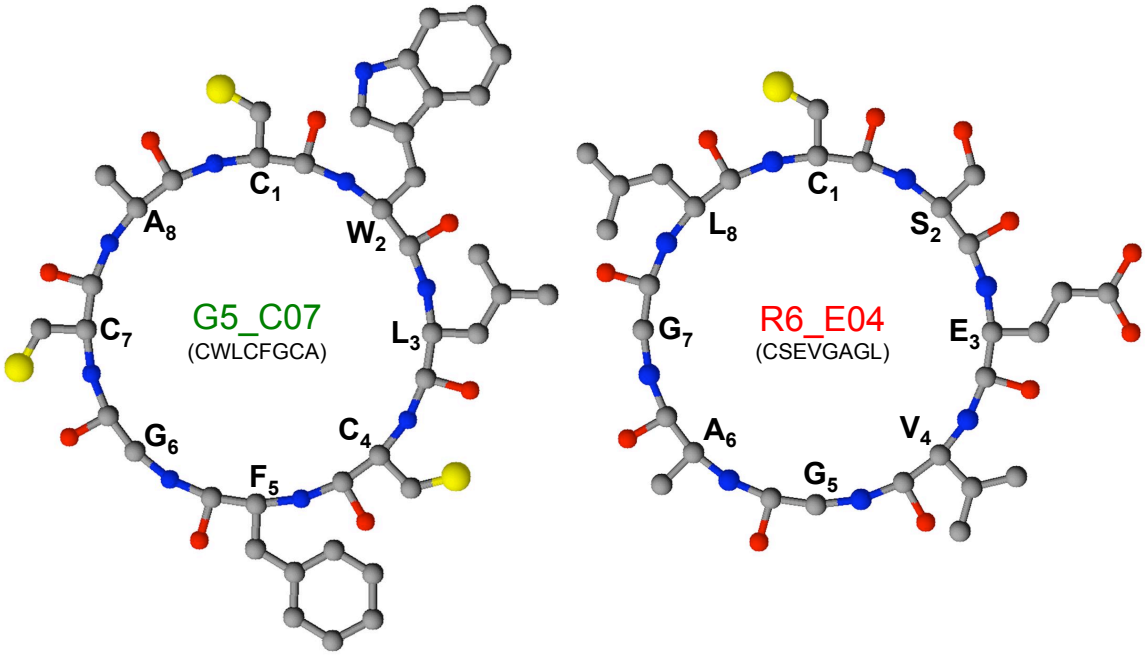


Illustration 3.3: Primary structure of CX7 cyclic peptide hits

Ball and stick amino acid structures for G5_C07 and R6_E04 are shown on the left and right respectively. The amino acids are labeled based on their primary sequence starting with the obligatory cysteine. Ball and stick models were constructed using ChemDoodle.

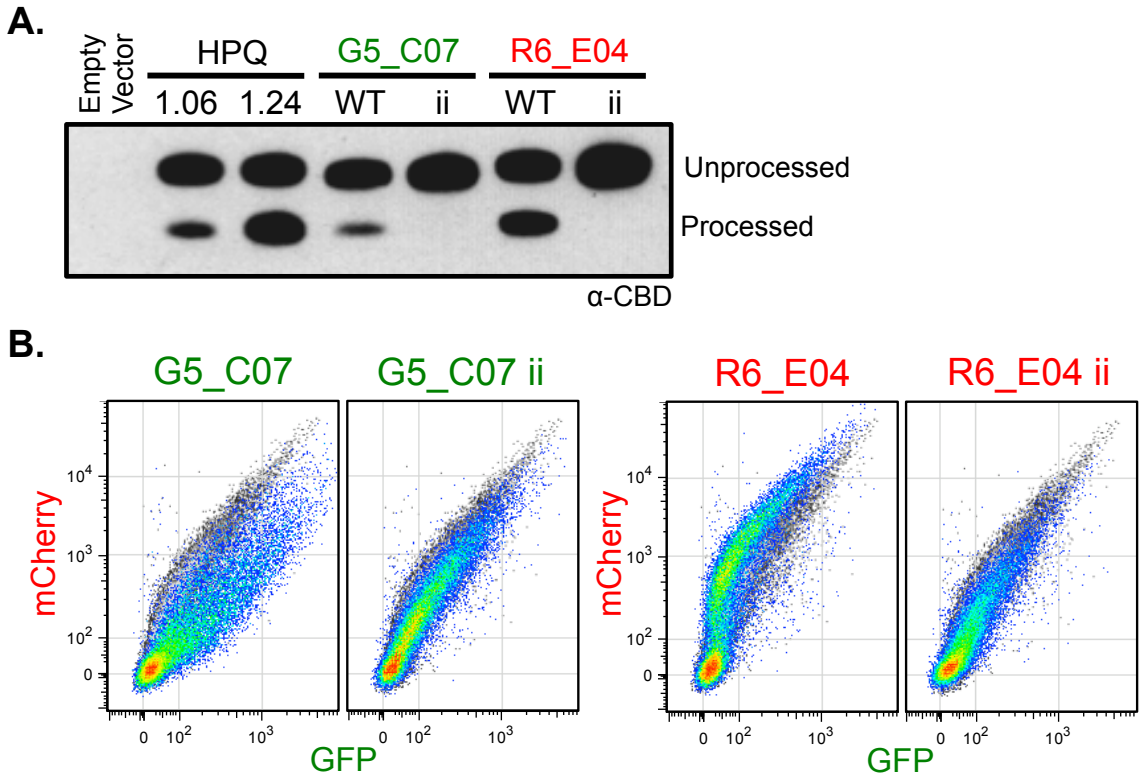


Figure 3.2: The reporter phenotypes of G5_C07 and R6_E04 are dependent on efficient cyclic peptide processing

(A) Western blot analysis of soluble protein from wild-type cells expressing HPQ, G5_C07, and R6_E04 cyclic peptide. For HPQ two OD (A_{600}) time points were taken and the OD is indicated above the respective lane. For G5_C07 and R6_E04 both the wild-type (WT) and intein inactivated mutation (ii; T69A H72A) are analyzed. (B) Reporter phenograph overlays with the indicated cyclic peptide construct (pseudocolor) overlaid on an empty vector control (grayscale).

In our western analysis of different HPQ samples (Figure 3.2A), it was clear that the processing of the control HPQ cyclic peptide is more efficient at higher ODs (A_{600}). To ascertain whether there are differences in reporter phenotype at different ODs, I performed a timecourse experiment in which I sampled a culture of yeast expressing either G5_C07 or R6_E04 at multiple time points spanning 10 hours. As with HPQ, the processing efficiency looks to increase at later ODs, which might have to do with decreased *de novo* expression when a culture nears stationary phase (Figure 3.3A). This experiment also reinforced our earlier observation that G5_C07 is processed less efficiently than R6_E04, even at later time points. When looking at reporter phenographs at selected time points, it is difficult to ascertain if there are significant differences, although time points around OD (A_{600}) look to have the most significant phenotype (Figure 3.3B). To look more closely at the individual reporter signals, histograms were compared for selected time points. For G5_C07 the greatest fluorescent levels in both channels are observed at OD (A_{600}) 0.83 (Figure 3.3C). In terms of R6_E04 we see a similar trend, except that the mCherry level is a log higher than G5_C07, with the highest levels of fluorescence at OD (A_{600}) 0.92. Together, these data show subtle differences across a growth curve in terms of cyclic peptide expression, processing, and reporter phenotype. For subsequent experiments I will analyze cells and cell material from cultures between OD (A_{600}) 0.8 and 1.0.

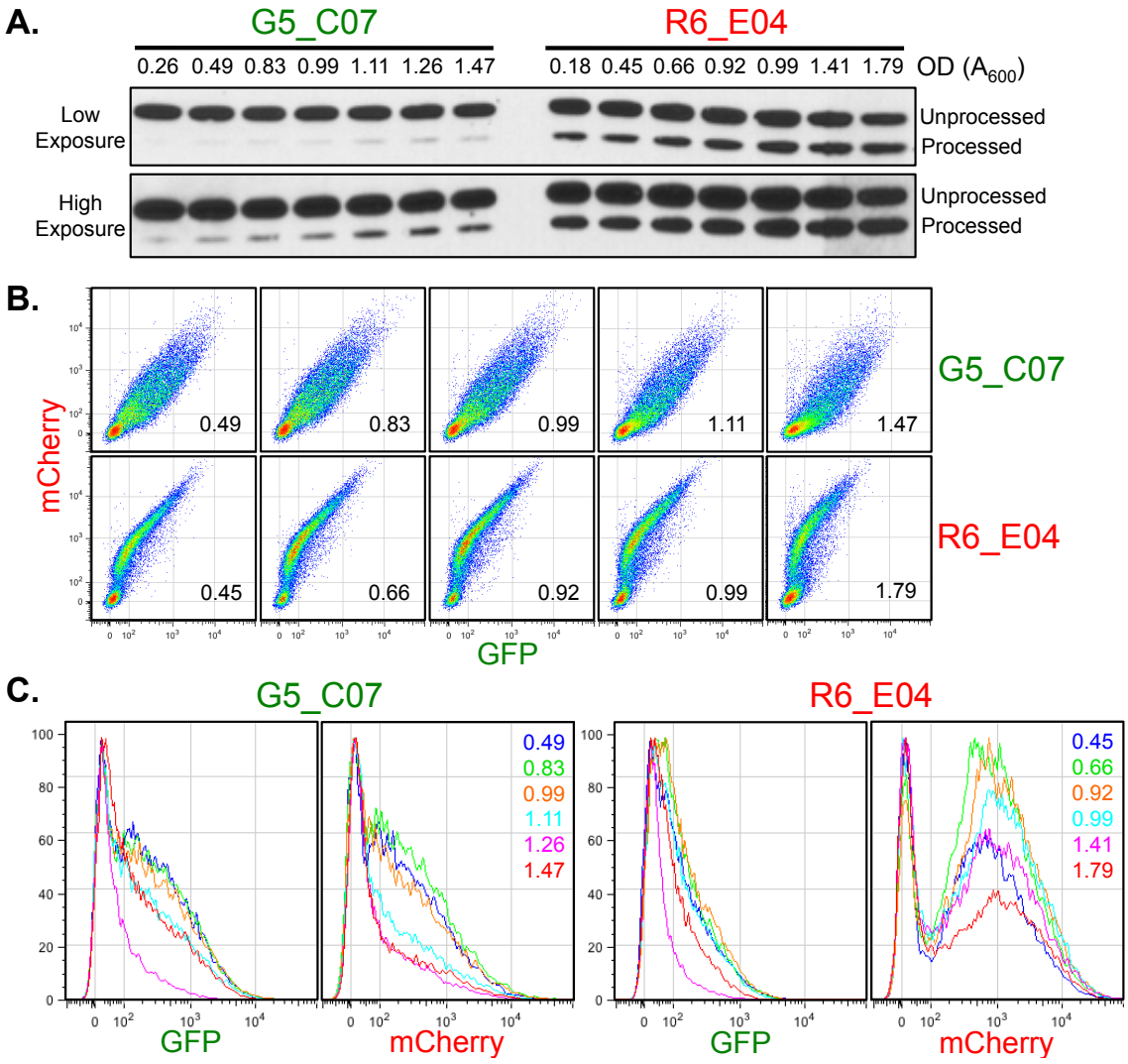


Figure 3.3: Timecourse of cyclic peptide expression, processing, and reporter phenotype

For all panels wild-type yeast harboring the gene expression reporter and either G5_C07 or R6_E04 were observed over the indicated ODs (A_{600}). Samples were taken at the indicated ODs for protein and flow cytometry. (A) Western blot analysis of protein collected from the indicated ODs, probing for the chitin-binding-domain on the cyclic peptide precursor. (B) Select reporter phenographs are shown for indicated time points of the cyclic peptide labeled to the right. (C) Histograms of flow cytometry data in order to compare the phenographs in panel B. Legend within the mCherry graph apply to GFP for the same cyclic peptide. The y-axis units are “% of maximum”.

3.3.2 Structure-activity relationship of G5_C07

In identifying two high-confidence cyclic peptide hits from our FACS guided screen, we subsequently desired to determine the role of the 8 amino acids within these peptides. Using site-directed PCR mutagenesis an alanine scan was performed; all non-alanine amino acids of G5_C07 (CWLCFGCA) were individually changed to alanine while the alanine was changed to a glycine. Mutation of the first cysteine was expected to eliminate processing due to its role in processing (Illustration 3.1C). The C1A mutation indeed abolished processing, but C1S was actually processed more efficiently that even WT (Figure 3.4A). In talking with our collaborators, this is not at all uncommon, since the serine can actually take part in the chemistry required for intein splicing. For the rest of the mutations, I noticed reduced levels of processing, except for C4A. Since most of these mutants resulted in decreased processing, I predicted that most of them would reduce the reporter phenotype. In comparing the phenographs from the mutants to both an HPQ control and WT G5_C07, it is quite obvious that C1, W2 are critical for the processing and activity of the cyclic peptide (Figure 3.4B). It is interesting that the C4A mutant is expressed and processed well, yet it eliminates the cyclic peptides, indicating that C4 is critical. Mutants F5A and G6A have intermediate affects on the activity of G5_C07 while L3A, C7A, and A8G have very little effect on the activity of the cyclic peptide. I mutated L3 and C7 to lysine to determine if this substitution can be used in linking synthesized peptide to an affinity matrix, which could then be used to identify cellular components that have

high affinity for the cyclic peptide. Both lysine mutations reduce the effect of the cyclic peptide on our reporter and most likely would not be an ideal substitution to use for target identification (Figure 3.4C). Although the L3K mutation retained some effect, this mutation decreases the overall expression of the cyclic peptide precursor to barely detectable levels (Figure 3.4D). A lysine substitution at position 8 was considered because the peptide still had a significant effect on reporter expression, but due to its close proximity to the critical C1 residue, we chose not to explore this route. The nullifying effects observed when mutating both C1 and C4 led me to believe that this cyclic peptide might contain a disulfide bond, which has been observed by our collaborators (Kritzer laboratory, data not published). Due to the low level of processing and the finicky behavior of this cyclic peptide, we decided to only further pursue R6_E04.

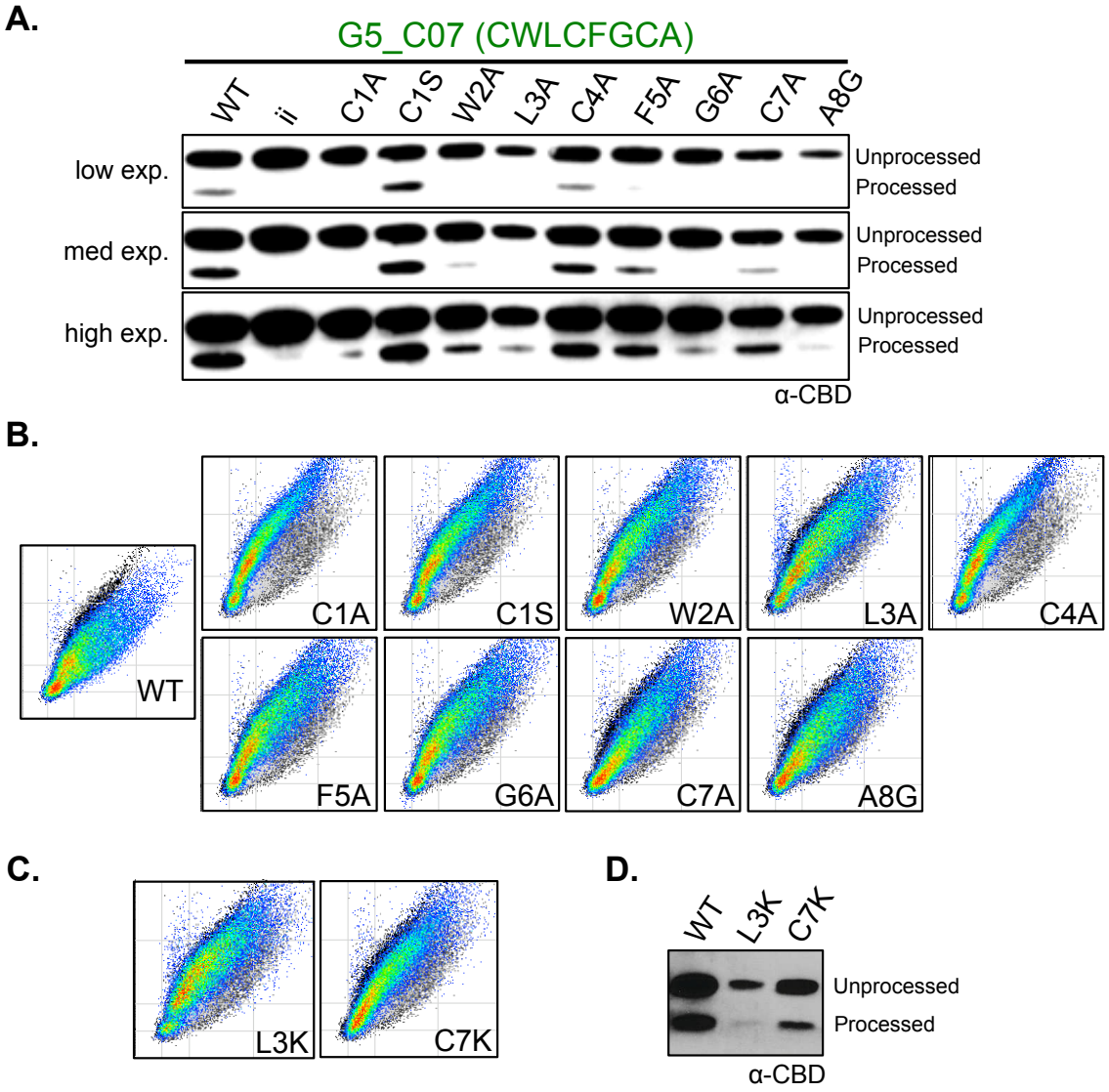


Figure 3.4: Alanine scan of G5_C07 determines most residues are critical to function

The CX7 cyclic peptide clone G5_C07 was altered by site-directed PCR mutagenesis. (A) Western blot analysis determining cyclic peptide expression and processing of wild-type (WT; CWLCFGCA), intein inactivated mutant (ii; T69A H72A), and the indicated cyclic peptides (labeled above lanes). Three exposures are shown. (B and C) Reporter phenograph overlays are shown with data from HPQ (negative control) cells in black, WT G5_C07 in grayscale, and the indicated clone in pseudocolor (except for the WT, in which it is only G5_C07 overlaid onto HPQ). (D) Western blot as in panel B analyzing mutants in panel C.

3.3.3 Structure-activity relationship of R6_E04

As our efforts with the green-shifted hit G5_C07 were inconclusive, I employed the same strategy for the red-shifted hit R6_E04 (CSEVGAGL). In contrast with the alanine scan mutations of G5_C07, those of R6_E04 appeared to express and process efficiently (Figure 3.5A). The only mutant with lower processing efficiency was C1A, which is surprising given that cysteine and serine side chains are typically used. It is possible that this mutant cleaves yet is not fully spliced. Although our collaborator has also observed this phenomenon, further experimentation would need to be done to understand the effects of a C1A mutation. In analyzing the effect of these alanine scan mutations on reporter expression, I noticed that as with G5_C07, almost all the substitutions resulted in decreased reporter shift when compared to the parental R6_E04 (Figure 3.5B). Of note, the C1A phenograph, which mimics wild-type conditions (no cyclic peptide), is consistent with a non-cyclic peptide-producing situation, as discussed above. The only mutants that resulted in significant reporter shifts were E3A and V4A, with V4A having a slightly more severe shift than the original cyclic peptide. I then substituted valine for lysine at position 4 and analyzed reporter expression (Figure 3.5C). To my delight the V4K clone not only retained the red-shift but also resulted in an even more severe shift than “wild-type” R6_E04 and was expressed and processed at similar levels (Figure 3.5D). These data led us to further explore using the fourth position in R6_E04 as a potential residue to use in target identification.

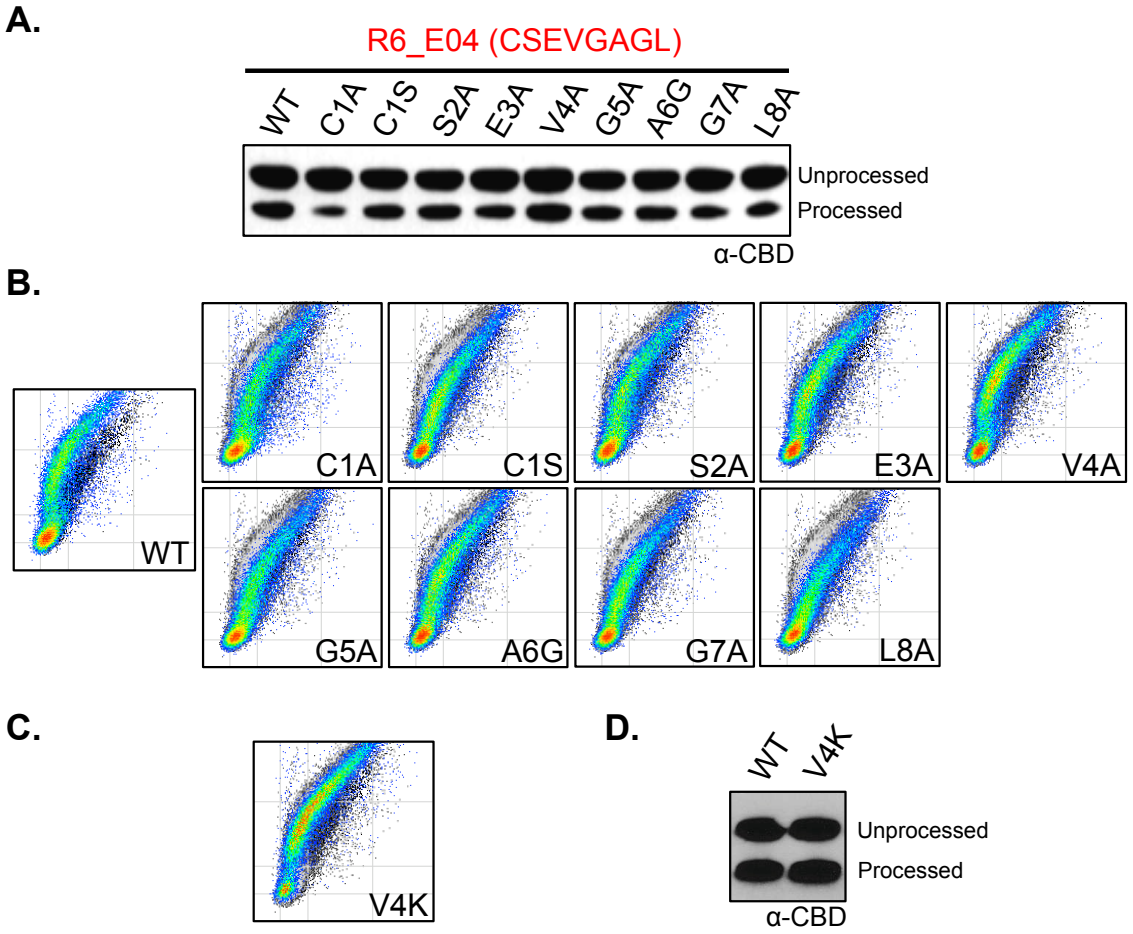


Figure 3.5: Mutagenesis of R6_E04 identifies critical and replaceable residues

Cyclic peptide clone R6_E04 was modified by site-directed PCR mutagenesis. (A) Western blot analysis determining expression and processing wild-type (WT; CSEVGAGL), intein inactivated mutant (ii; T69A H72A), and the indicated R6_E04 cyclic peptides (labeled above lanes). (B and C) Reporter phenograph overlays are shown with data from HPQ (negative control) cells in black, WT G5_C07 in grayscale, and the indicated clone in pseudocolor (except for the WT, in which it is only R6_E04 overlaid onto HPQ). (D) Western blot as in panel B analyzing mutant from panel C.

3.3.4 Development of a R6_E04 control peptide for biochemical target identification

After discovering the potential of using a lysine substitution at the fourth position of R6_E04 (CSEKGAGL), we wanted to make a cyclic peptide with the lysine in place that did not have any apparent activity. This peptide would then be used as a control in pull-down experiments. Our hypothesis was that if we eliminate essential residues of the R6_E04 cyclic peptide, they would have reduced affinity for the target of interest. In our earlier alanine screen, substitution of either the 3rd position glutamic acid (E) or 8th position leucine (L) with alanine muted the effect of the R6_E04 cyclic peptide (Figure 3.5B). Using site-directed mutagenesis, I replaced E3 (V4K control 1, CSAKGAGL), L8 (V4K control 3, CSEKGAGA), or both (V4K control 2, CSAKGAGA) with an alanine. These clones were expressed alongside our gene expression reporter and the resulting phenographs were analyzed (Figure 3.6A). All three control cyclic peptides displayed reduced activity when compared to the parental V4K clone (Figure 3.6B), resulting in a phenograph that is very similar to the negative control HPQ (Figure 3.6C). These findings were encouraging because I was looking to generate a R6_E04 V4K cyclic peptide with reduced or eliminated activity. Moreover, when I analyzed the expression and processing levels of these peptides, they all appeared to be expressed and processed well. Given these data, we chose to use R6_E04 V4K control 2 as a negative control for affinity chromatography.

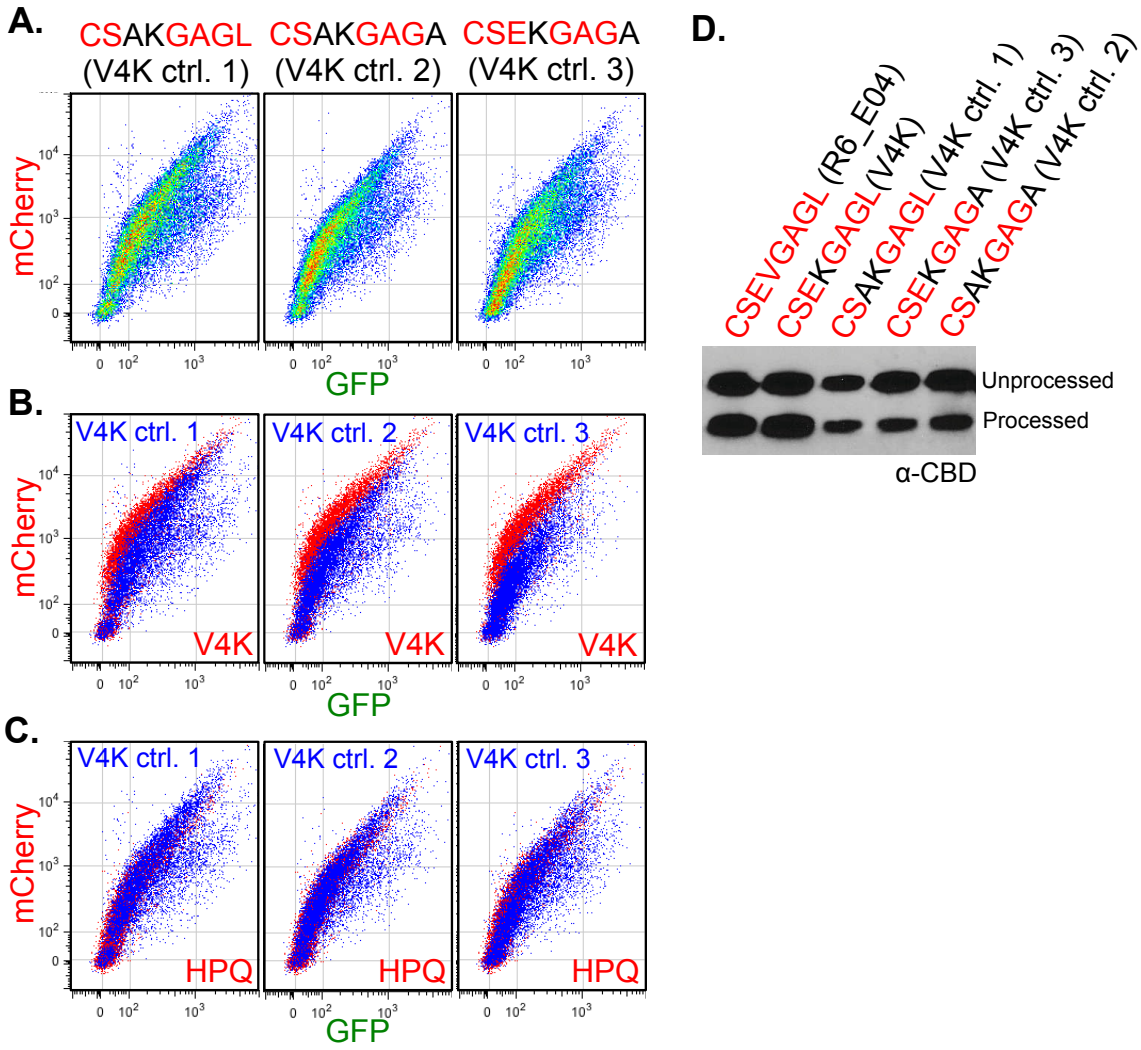


Figure 3.6: R6_E04 V4K control cyclic peptides mute reporter phenotype

(A) Reporter phenographs are shown for wild-type cells expressing the indicated R6_E04 V4K control (ctrl.) cyclic peptide construct (labeled above lanes). These are site-directed mutants of the V4K clone. To aid in comparison, these three phenographs are then overlaid onto a phenograph for the parental R6_E04 V4K (B) and the negative control cyclic peptide HPQ (C). The three V4K control cyclic peptides are in the foreground (blue) and the V4K or HPQ cyclic peptide data are in the background (red). (D) Western blot analysis determining expression and processing of the indicated R6_E04 cyclic peptides (labeled above lanes).

3.3.5 Pull-down attempts with biotinylated synthetic cyclic peptides

After determining that control 2 (CSAKGAGA) would be an attractive negative control for R6_E04 V4K (CSEKGAGL), we proceeded with a pull-down approach in an attempt to identify potential cellular targets. Our collaborators synthesized the aforementioned cyclic peptides with a biotin moiety linked to the lysine (Kritzer Laboratory, Tufts University). Cleared wild-type yeast extract was incubated with no cyclic peptide, cyclo-CSEKGAGL (experimental), cyclo-CSAKGAGA (control). Protein samples were analyzed by SDS-PAGE and a representative gel is shown in Figure 3.7. At first look, it is obvious that a number of endogenous proteins bind to streptavidin (Figure 3.7), which is not surprising given that budding yeast has a number of naturally biotinylated proteins (Hasslacher et al. 1993; Brewster et al. 1994; Kim et al. 2004; Hoja et al. 2004; Werven and Timmers 2006). Even with extensive clearing, there were detectable levels of proteins in the bound fraction when no biotinylated cyclic peptides were used (Figure 3.7). Unfortunately, we observed very little difference in the bound fraction from the pull-down with R6_E04 V4K (CSEKGAGL) or the control (CSAKGAGA; last lane). This experiment was repeated many times, while empirically changing conditions, yet I was unable to detect any specific proteins associated with the experimental R6_E04 V4K peptide. This being said, we did submit protein from the pull-downs for mass spectrometry analysis, however the results were inconclusive.

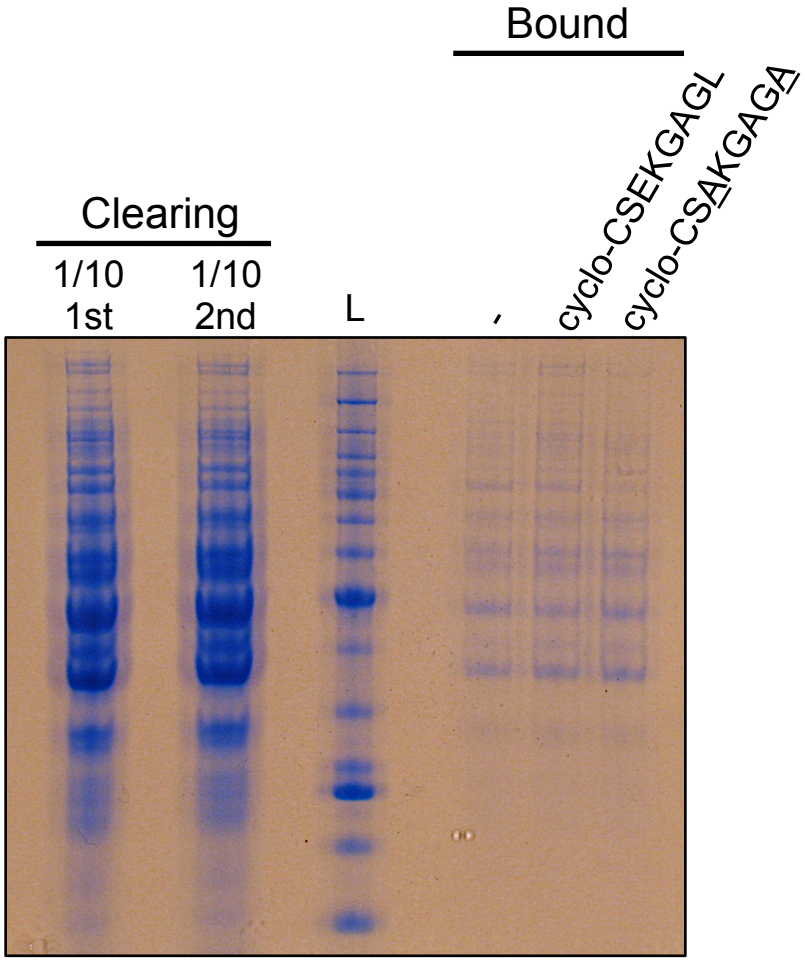


Figure 3.7: SDS-PAGE analysis of R6_E04 V4K pull-down

The indicated protein samples from a cyclic peptide pull-down experiment are shown after coomassie staining (L, Benchmark Protein Ladder; -, no cyclic peptide). For the clearing lanes, one tenth of the 1st and 2nd clearing was run on the gel. For the bound, half the material was analyzed by SDS-PAGE and the other half was saved for mass spectrometry analysis. The first cyclo-lane is the R6_E04 V4K peptide while the second (far right) is the control cyclic peptide.

3.3.6 The R6_E04 reporter phenograph mimics a pre-mRNA splicing defect

In conjunction with the biochemical efforts to identify the cellular target of R6_E04, I also utilized our existing reporter data to identify possible target pathways. Although it was quite obvious that the R6_E04 peptide elicited an effect similar to pre-mRNA splicing or nonsense-mediated decay mutants when looking at reporter behavior (Figure 2.3B-C), we wanted to compare the data in as unbiased a manner as is possible. To do so, I incorporated binned flow cytometry data from wild-type cells expressing R6_E04 with the data from the yeast deletion collection in Figure 2.14 and hierarchical clustering was used. Not surprisingly, all eight biological replicates of R6_E04 clustered within the red-shifted clade (Figure 3.8A). When focusing on the specific clade into which the R6_E04 samples grouped, I also observed the more severe non-essential pre-mRNA splicing mutants (*lea1Δ, ist3Δ*, and *bud13Δ*) in the same group (Figure 3.8B). When comparing the phenographs of these mutants and R6_E04, there is evident similarity (Figure 3.8C). The overall shapes of the phenographs are very similar, while R6_E04 cells penetrate the phenograph more. Additionally when considering the density of the cells in the pseudocolor plots, there is a noticeable increase in mCherry fluorescence. These higher levels of mCherry fluorescence are in-between that of the pre-mRNA splicing and nonsense-mediated decay mutants (Figure 3.8C).

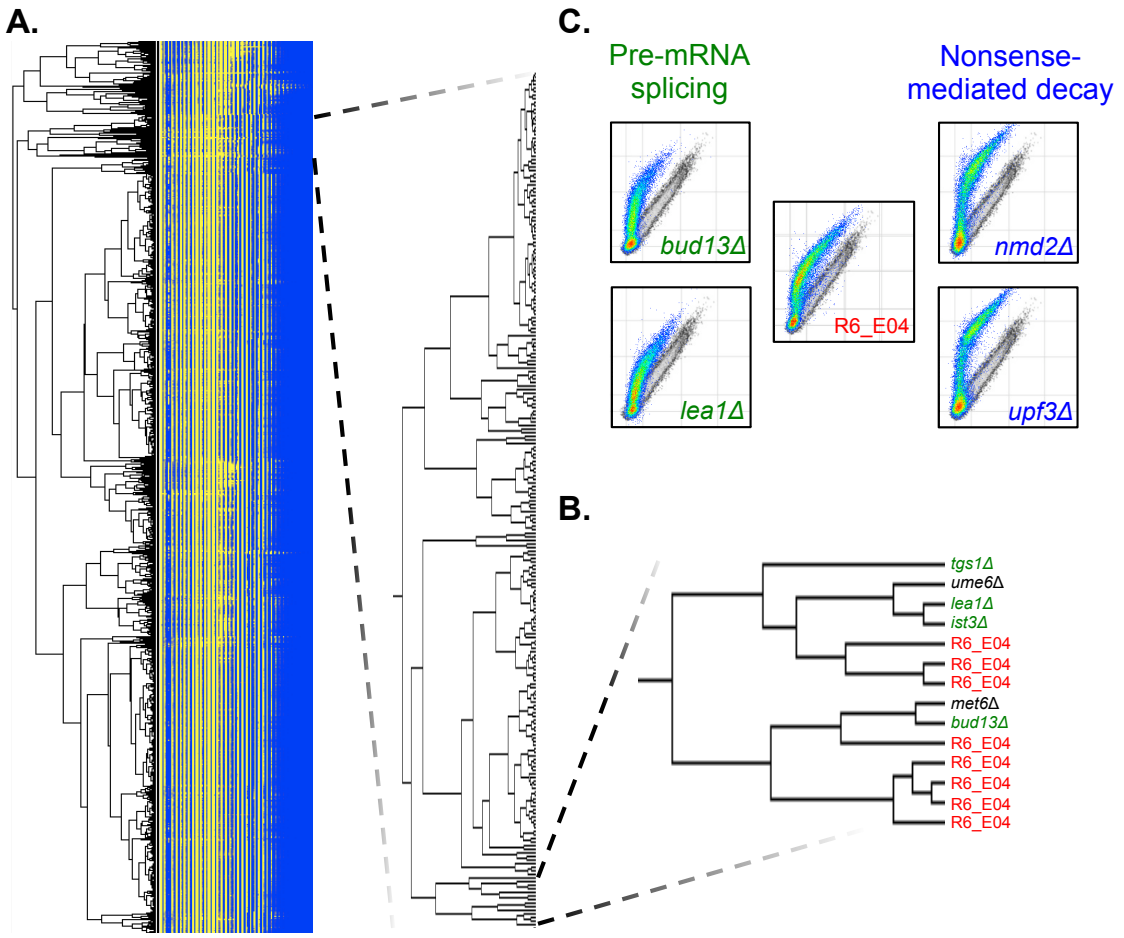


Figure 3.8: Binning and clustering analysis of reporter expression places R6_E04 with pre-mRNA splicing mutants

Flow cytometry data from wild-type yeast harboring the cyclic peptide R6_E04 and the gene expression reporter were binned and incorporated into the deletion collection dataset from Figure 2.14. The combined dataset was then clustered hierarchically (A). The “red-shifted clade” is enlarged to the right and the specific clade that contains all R6_E04 biological replicates is further expanded and annotated (B). Pre-mRNA splicing mutants and cyclic peptide samples are colored green and red respectively. (C) Representative phenographs are shown to aid in comparison. Phenographs are as usual with mCherry and GFP fluorescence plotted on the y- and x-axis respectively, with the same scales applied. The indicated mutant or cyclic peptide is in pseudocolor, overlaid onto wild-type cells in grayscale.

3.3.7 R6_E04 displays a minor gene-specific pre-mRNA splicing defect

Since yeast expressing R6_E04 clustered with yeast lacking certain pre-mRNA splicing factors (Figure 3.8), we wanted to determine if splicing of endogenous introns was impaired. To assay splicing efficiency, RT-qPCR was performed using total RNA isolated from cells expressing the cyclic peptides HPQ, R6_E04, R6_E04 ii (intein inactivated), and R6_E04 V4K. For two of the introns tested, *CYH2* and *RPL22B*, there was very little difference in splicing efficiency for the cyclic peptides tested (Figure 3.9). For the other two genes, *DBP2* and *TEF5*, I observed minor yet reproducible splicing defects for R6_E04 when compared to the ii control, as seen by an increase in the pre-mRNA/total ratio. It was concerning that I did not observe a splicing defect for the V4K construct, since it still retained its red-shifted reporter phenograph. Although this splicing defect of R6_E04 was reproducible, the relative fold increase of 1.5 and 1.4 is small compared to the 7 to 13 and 3 to 5 fold increases seen in pre-mRNA splicing mutants for *DBP2* and *TEF5* respectively (data not shown).

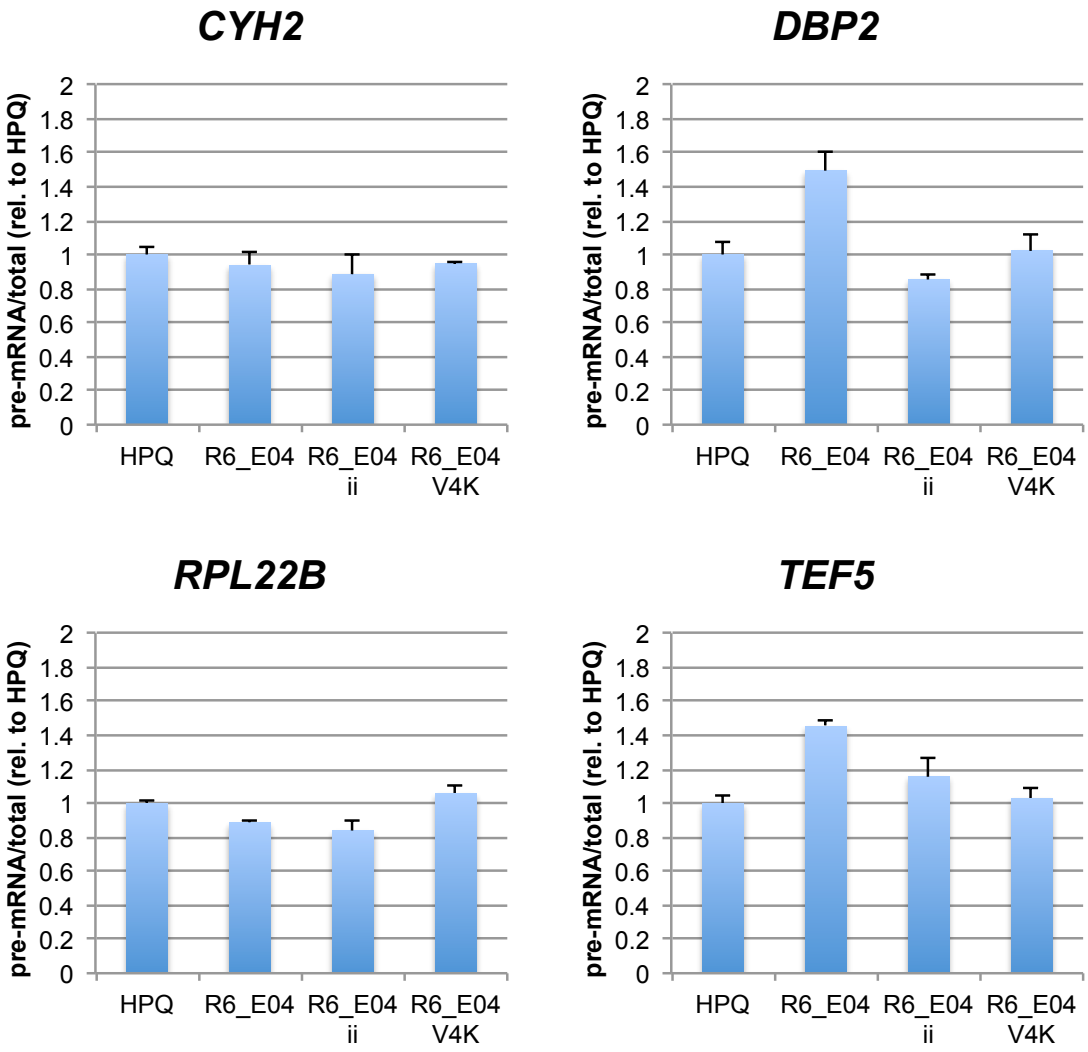


Figure 3.9: RT-qPCR reveals a minor pre-mRNA splicing defect in cells expressing R6_E04

RT-qPCR analysis was performed using total RNA from the indicated strains (ii, intein inactivated) for the specified gene. Primers for qPCR are described in Table 3.3. The quantities of pre-mRNA and total transcripts were quantified using a standard curve and ratios are relative to the HPQ negative control. Error bars represent the standard error between biological triplicates.

3.3.8 R6_E04 does not appear to inhibit the nonsense-mediated decay pathway

Given the similarity of the reporter phenograph of R6_E04 to mutants of the nonsense-mediated decay (NMD) pathway (Figure 3.8C), I tested whether cells expressing the R6_E04 cyclic peptide had elevated levels of *CYH2* pre-mRNA, a poorly spliced prototypical endogenous NMD substrate (He et al. 1993; He and Jacobson 1995). Included in this experiment were pre-mRNA splicing (*bud13Δ*) and NMD (*nmd2Δ*) controls. Northern blot analysis of total RNA for *CYH2* transcripts revealed no *CYH2* pre-mRNA accumulation in cells expressing R6_E04 or either of the cyclic peptide controls (HPQ and R6_E04; Figure 3.10). Since R6_E04 did not appear to have an NMD defect, we were curious if the reporter pre-mRNA was specifically stabilized, giving rise to the mCherry fluorescence increase. Because our gene expression reporter intron was based off *CYH2,* we were also able to detect reporter pre-mRNA levels with the same *CYH2* probe. In cells harboring our reporter and the cyclic peptide constructs, I did not observe any increase in reporter pre-mRNA levels (Figure 3.10). In all, these data do not indicate that the R6_E04 reporter phenotype is not due to inhibition of NMD.

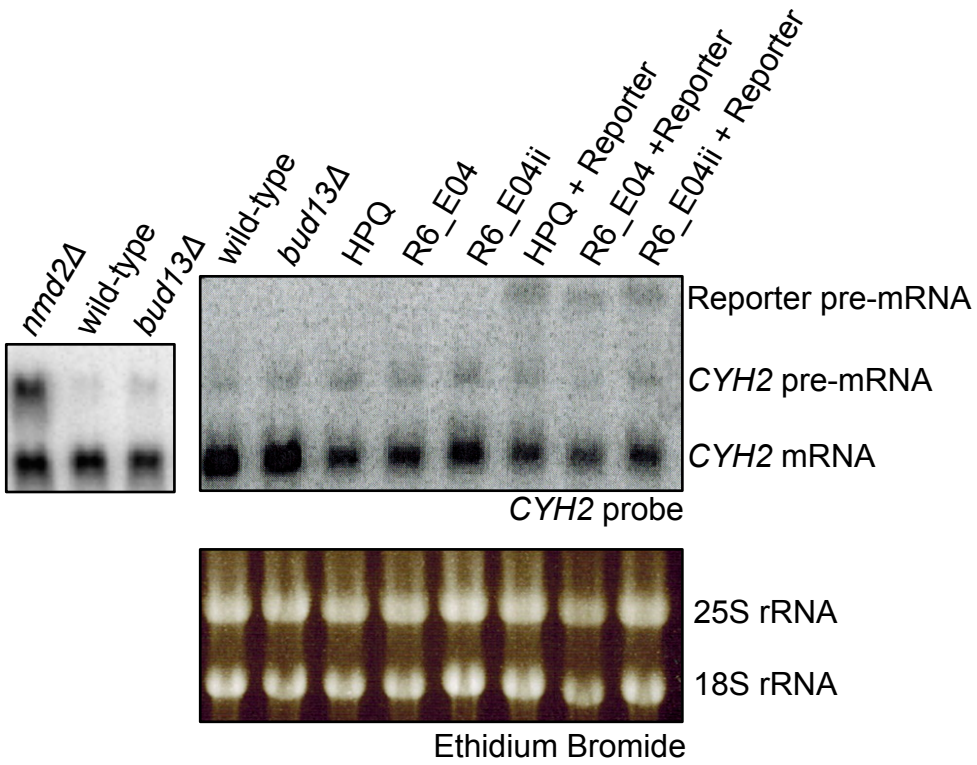


Figure 3.10: Northern blot analysis reveals no nonsense-mediate decay defect in cells expressing R6_E04

Total RNA was electrophoresed in an agarose gel, visualized for rRNA (ethidium bromide; lower panel), transferred to nylon, then probed for reporter pre-mRNA and *CYH2* transcripts with a body-labeled random-primed *CYH2* probe. Transcript species are labeled to the right and apply to both blots. Lanes are labeled with the cells the RNA was isolated from, make note that the three right-most samples harbor our gene expression reporter (pRS415 *TDH3*-Reporter; MS169) in addition to the cyclic peptide constructs. The ethidium stained 18S and 25S rRNA are used to verify RNA loading levels.

3.3.9 Comparison of plate reader analysis with flow cytometry for measuring reporter expression

In almost all the experiments performed with our gene expression reporter, flow cytometry analysis was used to measure the expression of GFP and mCherry. In preparing for high-throughput drug screen, it would be desirable to complete the primary screen using a plate reader since it was possible to analyze a 96-well plate in a matter of seconds, compared to 20 minutes with the flow cytometer. Instead of analyzing the fluorescence of individual cells, a plate reader measures the bulk fluorescence within a given well. Since mutations in specific gene expression processes resulted in characteristic flow cytometry phenographs, it was critical to determine if I could detect these differences with the plate reader. I used a set of prototypical deletion strains with the standard *TDH3*-Reporter. When the cells were ready for analysis, the samples were split in half, enabling analysis by both flow cytometry and plate reader (Figure 3.11). We were encouraged to see that the pre-mRNA splicing and NMD mutants (*bud13Δ* and *lea1Δ, upf1Δ* respectively), which have a low GFP/mCh ratio, were significantly separated from wild-type. On the other end of the spectrum the mRNA export mutant *sac3Δ,* which results in increased levels of GFP and decreased levels of mCherry, was also easily identifiable. With this analysis in hand, we were confident that we would be able to detect defects in gene expression with our gene expression reporter and a plate reader.

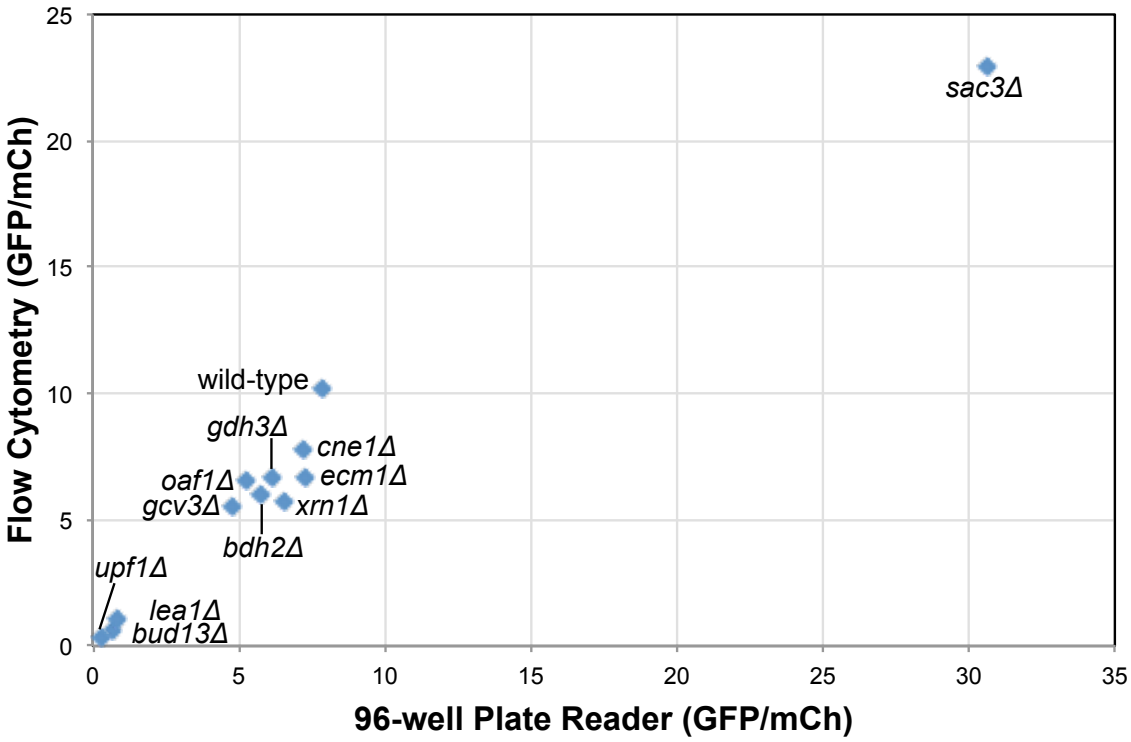


Figure 3.11: Reporter expression data generated from flow cytometry and plate reader are comparable

GFP/mCherry ratios were calculated by flow cytometry (single-cell) and plate reader (bulk) analysis and plotted. The gene deletions are indicated with labels.

3.3.10 Validation and optimization of GM ABC-16 for drug screening with a copper inducible reporter

In preparing our reporter system for high-throughput screening with Ti-3D (Texas Institute for Drug & Diagnostic Development, The University of Texas at Austin), I first needed to determine if there was appreciable levels of GFP fluorescence from the drug-sensitive yeast strain GM ABC-16 since the creators of the strain used a doxycycline inducible GFP system in the creation of the strain (Suzuki et al. 2011). Since there were 16 copies of an inducible GFP open reading frame in this strain, it was possible there would be background GFP fluorescence that could interfere with the GFP signal resulting from our reporter. In Figure 3.12A I demonstrate that there is negligible GFP fluorescence when compared to our standard wild-type yeast strain (BY4742; no GFP), confirming that this strain was suitable for our reporter assay. I then assayed reporter expression levels from the pRS413 *CUP1*-Reporter in the same strains to confirm that the 16 gene deletions in GM ABC-16 did not affect reporter expression. In Figure 3.12B I show that the copper inducible reporter expresses at similar levels in the drug-sensitive GM ABC-16 strain than our standard wild-type strain, further confirming the potential of this strain for our assay.

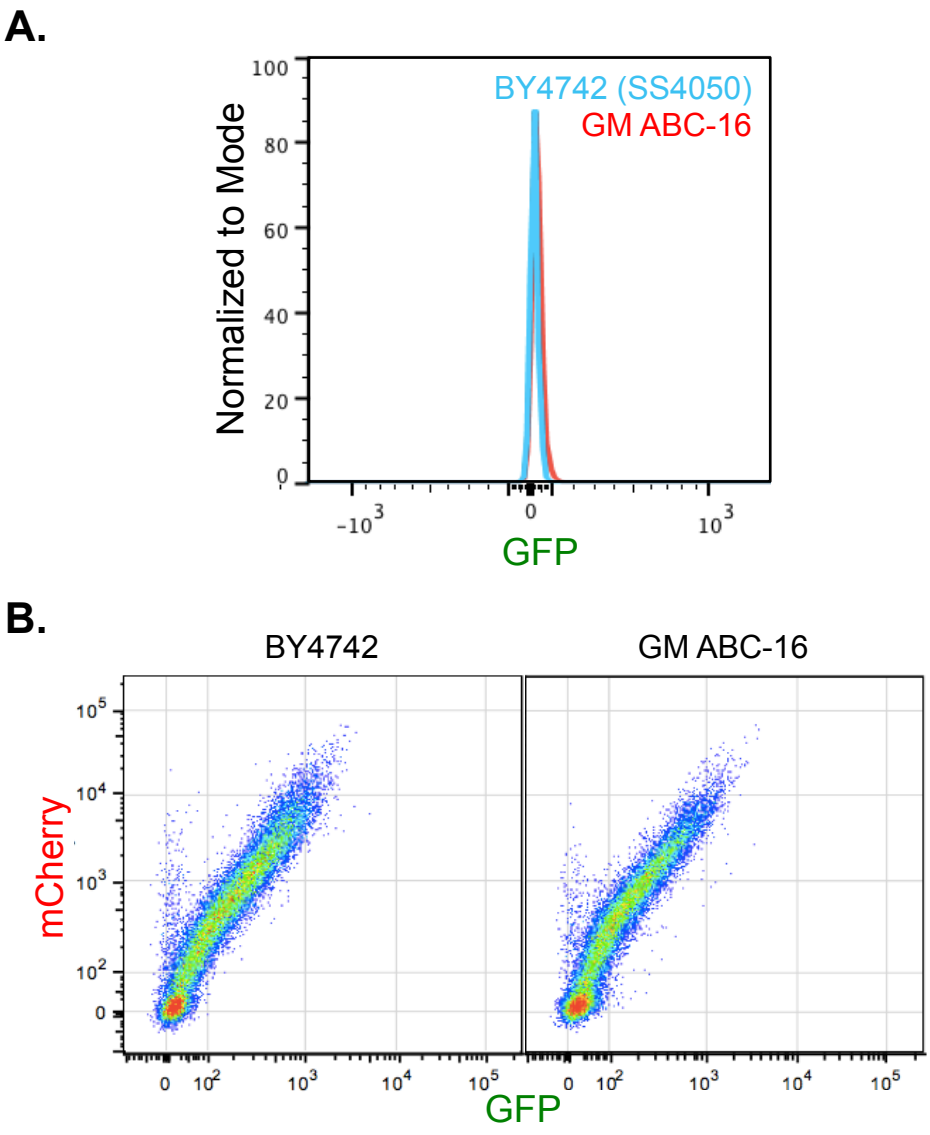


Figure 3.12: The drug-sensitive yeast strain GM ABC-16 is suitable for drug screen with inducible reporter

(A) Yeast strain GM ABC-16 (SS4127) has no appreciable GFP fluorescence compared with the standard WT yeast strain BY47472 (SS4050) when comparing GFP levels on a histogram. (B) Strain GM ABC-16 expresses comparable levels of GFP and mCherry from the pRS413 *CUP1*-Reporter in the presence of 1 mM $CuSO_4$ for 6 hours.

Our colleagues at Ti-3D challenged us to determine the conditions in which our reporter expressed the highest level of signal in the shortest period of time, in order that there was a suitable signal to measure. Early in our validation attempts, we determined that standard synthetic dropout media had significant autofluorescence, which is a result of riboflavin and folic acid. We began using low-fluorescent synthetic dropout media in all drug-screening efforts. We empirically determined that our drug screening strain, GM ABC-16 + pRS413 *CUP1*-Reporter, was optimally induced when mid/late exponential phase cells (OD (A_{600}) of 0.9-1.1) were diluted to early/mid exponential phase (OD (A_{600}) of 0.25) and treated with 0.5 mM $CuSO_4$ for 4-6 hours. Although we observed an increasing level of fluorescence after 6 hours, the cells were reaching the end of the exponential growth phase, and we wanted to analyze the cells while they were actively growing.

During the course of assay validation, we desired to use a control compound that permitted cells to grow, yet inhibit expression of our reporter, or at least shift the phenograph significantly. This compound would be used as a plate-to-plate control, as well as to test the robustness or our assay. After searching for compounds, we decided to use the absence of $CuSO_4$ as a "positive" control, which results in 100% inhibition of GFP and mCherry fluorescence. Although we would have liked to use a chemical compound as a positive control, we proceeded with these controls in our validation efforts. To determine the effect size of our assay, the Z-prime (Z') score was calculated,

which takes into consideration the mean and standard deviation of 48 control replicates (half a 96-well plate for each control), and difference between the negative and positive control (Figure 3.13, see Materials and Methods 3.2.13 for equation, parameter definition, and explanation). The Z' was calculated based on the fluorescence reading divided by the OD (A_{600}), which normalized the wells based on cell density. Both the GFP and mCherry Z' values, 0.82 and 0.85 respectively, were excellent. This indicated that our assay had great reproducibility across the plate and sufficient difference between the negative and positive controls. This experiment was repeated two additional times and yielded similar values, providing us great confidence in proceeding with our screening efforts.

GFP/OD Values

	1	2	3	4	5	6	7	8	9	10	11	12
A	21354	7028	21412	6937	21115	6836	20733	6542	20367	6547	20398	6405
B	20922	6603	20715	6820	21782	6797	20480	6623	20766	6348	20576	6333
C	20866	6913	19976	6908	20651	6685	20321	6572	20122	6397	20101	6441
D	20749	7063	20401	7194	19759	6733	20274	6667	21555	6431	20018	6349
E	20852	6883	20282	6860	20314	6719	20377	6679	20118	6351	19882	6511
F	21278	6952	20637	7002	21210	6826	20612	6562	19499	6491	19906	6496
G	21260	7019	20918	7104	20767	6768	20705	6775	20487	6628	19811	6606
H	21302	7235	20731	6864	21106	6814	20691	6781	22621	6648	20163	6585

σ_p = 584.27
σ_n = 234.04
μ_p = 20644.61
μ_n = 6715.30

Z' = 0.82

mCherry/OD Values

	1	2	3	4	5	6	7	8	9	10	11	12
A	6254	85	6277	58	6102	46	6122	52	5874	51	5956	62
B	6306	68	5990	46	5992	51	5811	70	5850	46	5710	30
C	6199	49	5433	56	5983	51	6034	63	5936	67	5752	69
D	6477	57	5987	15	6178	89	5952	69	6132	59	5655	53
E	6479	74	6096	64	6195	23	6098	56	6068	57	5646	47
F	6456	54	6209	63	6307	53	6168	58	6206	40	6000	56
G	6429	65	6461	50	6237	53	6373	52	6125	60	5884	42
H	6839	65	6322	70	6566	72	6434	52	6683	56	5876	54

σ_p = 281.29
σ_n = 13.41
μ_p = 6127.47
μ_n = 56.19

Z' = 0.85

Negative Control (Copper + DMSO)

Positive Control (DMSO)

Figure 3.13: The 96-well plate adapted reporter assay is robust and suitable for high-throughput screening

The fluorescence/OD (A_{600}) ratios were calculated for GFP and mCherry for a 96-well plate containing 48-wells of negative control (blue) and 48-wells of positive control cells. GM ABC-16 cells with pRS413 *CUP1*-Reporter were cultured in low-fluorescent synthetic dropout media lacking histidine to late exponential phase. Cells were diluted to OD (A_{600}) 0.25 and distributed across the plate and 0.5 mM $CuSO_4$ was added to negative control wells and the plate was incubated for 6 hours (via robotic pipetting). The Z-prime (Z') was calculated as in the Materials and Methods 3.2.13 (σ_p = standard deviation of positive control, σ_n = standard deviation of negative control, μ_p = mean of positive control, μ_n = mean of negative control). This experiment was repeated twice, resulting in similar Z' values.

3.3.11 Initial screens of a diverse set of 3647 small molecules

Through communication with our Ti-3D collaborators, we decided to begin our high-throughput small molecule screening efforts with a diverse set of small molecules. The first chemical libraries we screened were the NIH Clinical Collections (Molecular Libraries Small Molecule Repository operated by Evotec). This library consists of 727 small molecules that have been used in clinical trials, resulting in a collection with relatively high bioactivity and therapeutic potential. The next library tested was an institutionally derived collection of 600 kinase inhibitors. Kinase inhibitors are a very common class of therapeutic and biologically useful compounds (Chen 2012; Harrison 2012; Fabbro et al. 2012). The last and most diverse set of small molecules screened was The Spectrum Collection (MicroSource). This collection contained 2320 compounds and serves as a great primary screening tool due to its wide range of biological activity and structural diversity. 60% of this library is comprised of existing drugs, 25% are natural products, and the remaining 15% are “other bioactive compounds” such as non-drug toxins, enzyme inhibitors, membrane active compounds, and receptor blockers. Together, these three diverse collections of small molecules provided an excellent starting point for our screening efforts.

Our screening strategy is outlined in Illustration 3.2 and in our primary screen we analyzed the aforementioned libraries at 10 μM by plate reader analysis with the drug-sensitive yeast strain GM ABC-16. OD (A_{600}), GFP fluorescence, and mCherry fluorescence were measured for all wells. OD (A_{600})

normalized fluorescence values were used to calculate % inhibition, relative to the negative control samples. Compounds were ranked based on the % inhibition of GFP and mCherry separately and compounds with inhibitions of more than 50% were chosen for a secondary screen. Additionally, cytostatic compounds, as seen by no increase in OD (A_{600}), were selected, since they might reveal an interesting phenotype at lower concentrations. In all, 354 small molecules were selected to be screened in a secondary screen. The secondary screen was performed in the same manner as the primary screen, except that three concentrations were tested (10 μM, 1 μM, and 100 nM). Additionally, high-throughput flow cytometry analysis was used to examine reporter fluorescence in addition to plate reader analysis. The three concentrations served to determine if the compounds resulted in dose-dependent effects, as well as to gauge the effect of compounds that were toxic at the primary concentration 10 μM.

3.3.12 The secondary small molecule screen resulted in 16 high-confidence hits

Of the 354 small molecules that were analyzed in the secondary screen, we identified 16 compounds with significant dose-dependent effects on reporter expression. In Figure 3.14 I display the phenographs from the secondary screen for 10 of the 16 compounds. Across the board, the dose-dependent responses are apparent. It was also very encouraging that a number of the compounds still have a significant reporter phenotype at the low concentration of 100 nM.

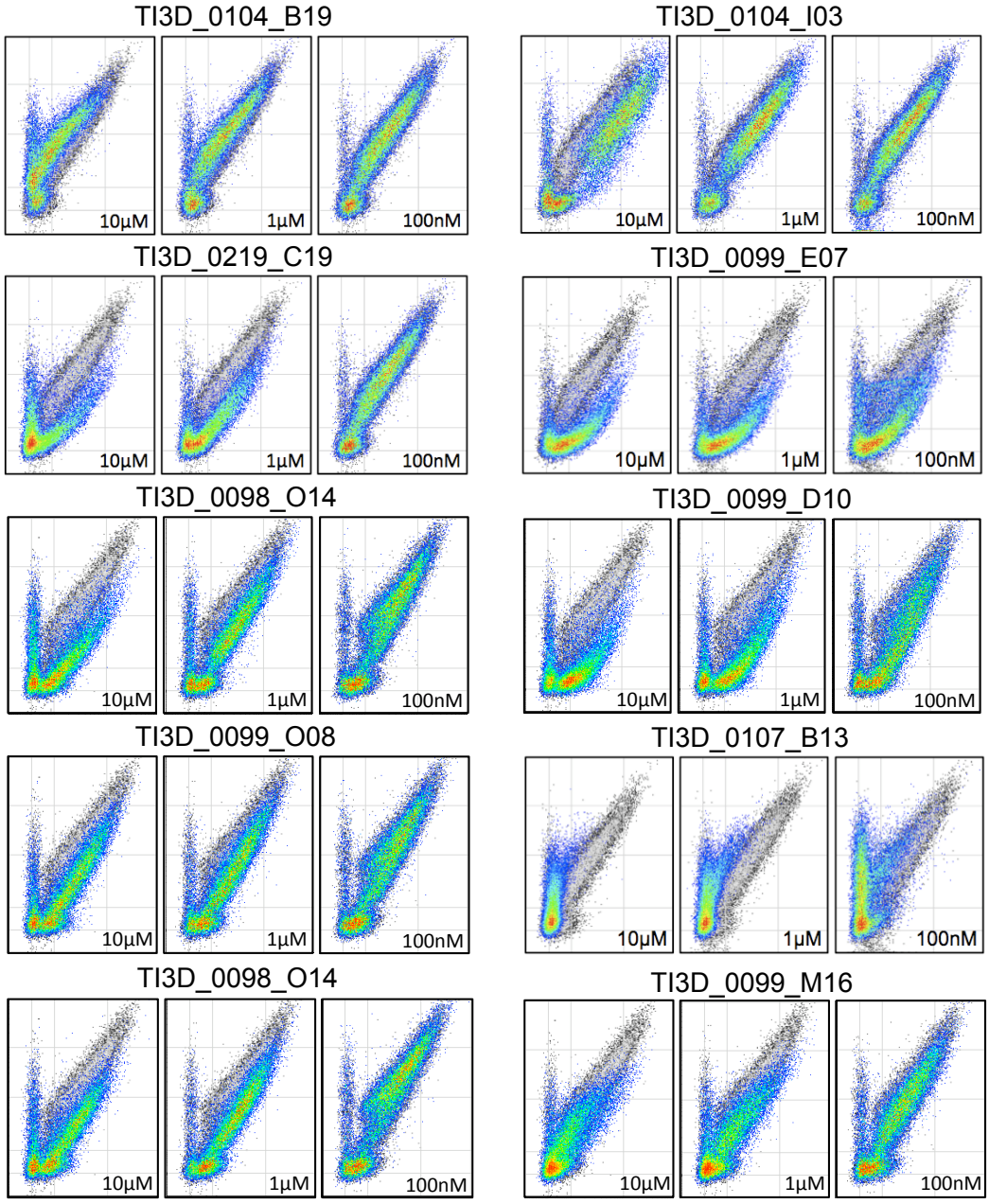


Figure 3.14: Dose-dependent reporter effects by high-confidence hits from small molecule screen

Small molecule treated flow cytometry data (pseudocolor) are overlaid onto DMSO only data (grayscale) for the three concentrations tested in the secondary screen. Phenographs are as usual with GFP and mCherry displayed on the x- and y-axis respectively

TI3D_0099_E07 was an interesting hit because there were three structural derivatives that resulted in almost identical reporter phenotypes. These three analogs of TI3D_0099_E07 only differ by one R group and represent a family of inhibitors for a specific kinase. Identifying four very similar small molecules with analogous phenotypes was very encouraging and results in higher confidence in the validity of this hit. Another interesting group of small molecules were ones that result in a phenotype alike TI3D_0104_I03. There were a total of three molecules with a more heterogeneous green-shifted phenotype, with the bulk of cells separated from the non-fluorescent cells at the origin. When investigating the nature of these compounds, all three were nucleoside analogs.

Our Ti-3D collaborators then attempted to determine the half maximal effective concentration (EC_{50}) for the hit compounds, which essentially is the concentration that elicits a response halfway between the negative control and the maximal response. EC_{50} values were calculated for each fluorescent signal separately, as well as OD (A_{600}; Figure 3.15). To complete these experiments fresh compounds were purchased in order to confirm the validity of the hit, as well as make a range of stock concentrations. The concentrations tested varied between compounds based on the severity of the reporter phenotype and growth inhibition, but all spanned at least a 4-log dilution series. We also determined the EC_{50} for wild-type yeast (BY4742/SS4050). Although a wide range of concentrations was used for each compound, some EC_{50} values were not able to be determined.

When studying the EC_{50} values, we were very encouraged by the results of some compounds at very low concentrations, including a number in the low to sub-micromolar range. We also observed overall lower EC_{50} values for the drug-sensitive strain GM ABC-16 when compared to wild-type yeast, which was surprising. Moreover, it was very interesting to observe some compounds with significantly different EC_{50} values for the same compound when comparing the effect on GFP and mCherry. For example, the mCherry EC_{50} for TI3D_0098_014 is nearly 10-fold lower than that of GFP. On the other hand, the GFP EC_{50} for TI3D_0104_I03 is 20-times lower than the mCherry EC_{50}. These differences reveal specific effects on reporter expression instead of an overall decrease in expression.

	Wild-type			GM ABC-16		
	Abs	GFP	mCherry	Abs	GFP	mCherry
TI3D_0098_O14	-	-	-	4.35	4.58	0.501
TI3D_0099_D10	2.58	-	0.54	0.0365	0.843	0.0355
TI3D_0099_E07	0.126	0.0363	0.0209	0.0268	0.0589	0.0415
TI3D_0099_I03	12.9	13.3	14.3	12.9	24.0	32.6
TI3D_0099_M16	0.0538	0.0191	0.0255	0.0467	0.0101	0.0278
TI3D_0099_O08	-	179	48.9	27.4	6.93	3.57
TI3D_0100_H19	0.036	0.184	0.0227	0.0144	0.0371	0.0122
TI3D_0104_A05	25.6	8.3	7.3	17.4	9	9.12
TI3D_0104_B19	18.1	-	-	6.14	-	9.77
TI3D_0104_I03	40.6	10.7	-	3.42	0.553	12.8
TI3D_0104_I09	13.7	55	7.15	8.9	5.4	3.70
TI3D_0105_C07	40.2	2.45	5.56	35	2.68	12.6
TI3D_0107_B13	4.1	1.35	1.26	0.486	0.372	0.101
TI3D_0219_C19	9.94	12.5	5.07	1.41	2.23	0.489

Figure 3.15: EC_{50} values for high-confidence small molecule hits

The EC_{50} values for high-confidence small molecule hits are shown. Our Ti-3D collaborators determined values by screening compounds across a concentration gradient of 6 concentrations covering at least 4-logs. All values are in μM and represent the concentration of compounds that elicits 50% of the maximal effect. Wild-type cells are strain SS4050 and GM ABC-16 are strain SS4127. Abs represents absorbance (OD (A_{600})) while GFP and mCherry refer to fluorescence. The "-" means the EC_{50} was unable to be determined.

3.3.13 Identification of possible target pathways for small molecule hits

As with the cyclic peptide hits, we wanted to identify the cellular targets of the small molecules that we identified as having potential effects on gene expression. The first attempt we made at accomplishing this goal was to use our binning and clustering pipeline, CHUBACA, to identify phenographs that are similar to deletion mutants. Our prediction was that a small molecule inhibiting a specific gene expression process would mimic a genetic perturbation of the same process. I binned and clustered the small molecule data from the high confidence hits from all three concentrations tested, and then integrated them with several dozen gene expression related deletion mutants. The resulting dendrogram is shown in Figure 3.16A. Of note, I observed all the small molecules that resulted in a green-shift clustering with the deletion mutant *rrp6Δ*, the 3' to 5' exonuclease of the nuclear exosome. Although none of the compound-treated phenographs are as severely shifted as the *rrp6Δ* mutant, they are all intermediates between the DMSO only control and *rrp6Δ* (Figure 3.16B). Additionally, one of the red-shifted hits, TI3D_0104_B19, clustered specifically with the splicing mutant *bud13Δ* (Figure 3.16A top right).

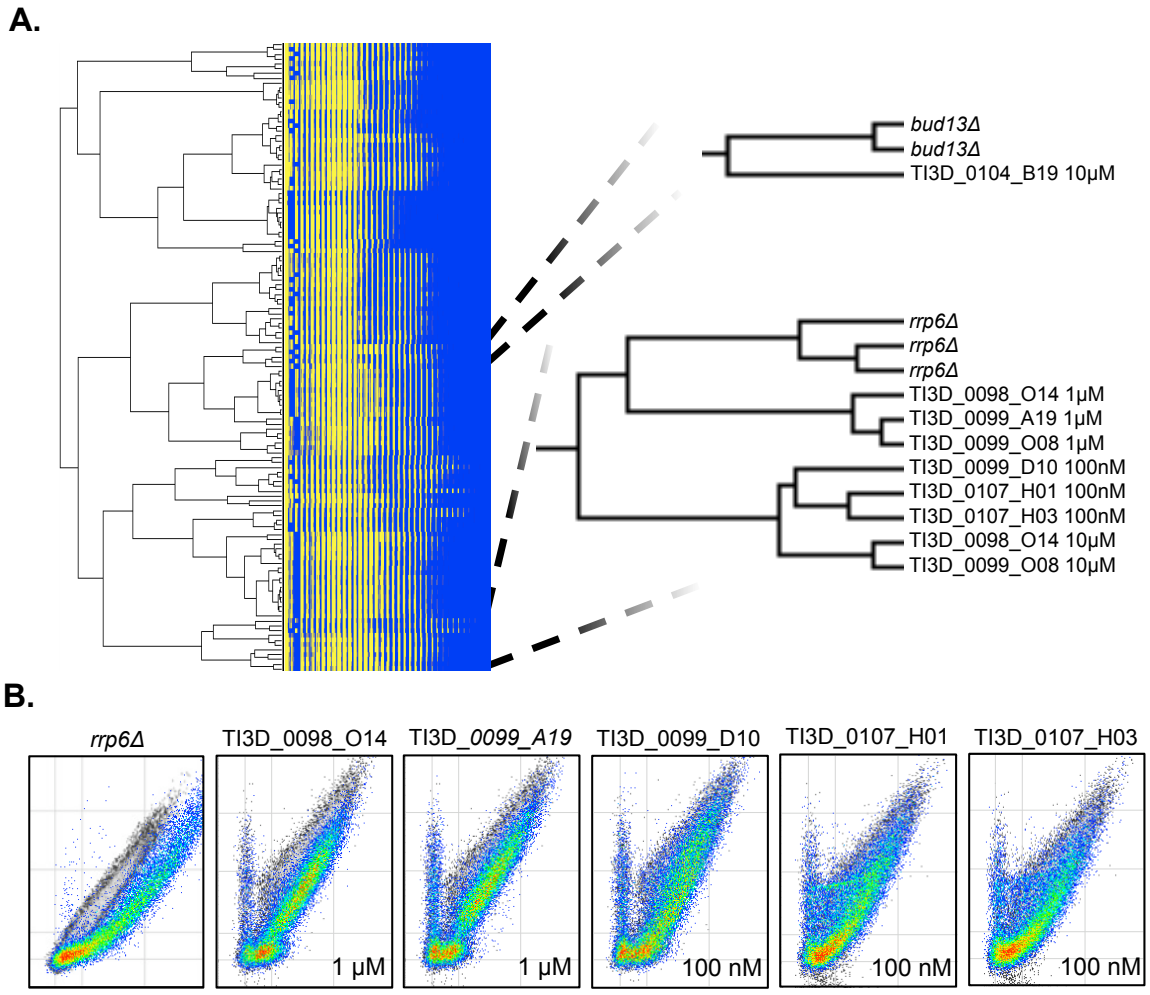


Figure 3.16: Preliminary drug target pathways revealed by clustering with deletion mutants

(A) A dendrogram depicting the similarity of flow cytometry phenographs is shown. Flow cytometry data from small molecule-treated cells were analyzed with a data set of deletion strains corresponding to the gene expression machinery. Each node represents a single data set and clades of interest are enlarged and annotated to the right. (B) Gene expression reporter phenographs representing data in the *rrp6Δ* green-shifted clade (enlarged in panel A). Phenographs are as usual with GFP and mCherry displayed on the x- and y-axis respectively, with the drug-treated cells in pseudocolor and negative control cells (DMSO only) in grayscale.

To go one step further in identifying the potential target pathways for the hit compounds, I employed RT-qPCR analysis to look at the level of endogenous transcripts. The first transcript analyzed was the *RPL18B* pre-mRNA, a known *in vivo* substrate of Rrp6p (Sayani and Chanfreau 2012). Since many compounds clustered with the *rrp6Δ* mutant strain, this was an obvious start to determine if Rrp6p might be inhibited. In this analysis I also included a control, 5-fluorouracil (5-FU), which has been shown to inhibit Rrp6p (Fang et al. 2004; Kammler et al. 2008), although the mechanism of action is unknown and most likely indirect. *RPL18B* pre-mRNA levels were normalized to both the Pol II transcribed *TDH3* and the Pol III transcribed *SCR1*. The concentration of compounds was designed to reflect the EC_{50} value. The highest levels of *RPL18B* pre-mRNA were detected in the 5-FU treated cells. Not surprising, some of the hit small molecules had elevated levels of *RPL18B* pre-mRNA, including TI3D_0219_C19, TI3D_0104_I03, TI3D_0105_O17, and TI3D_0107_B13. The first three of these compounds are green-shifted, similar to 5-FU, and two have been shown to inhibit Pol II transcription. I believe that transcription inhibition by these compounds leads to increased aberrant transcripts. This would overwhelm Rrp6p and lead to decreased processing efficiency, thus elevating the levels of its substrate *RPL18B* pre-mRNA. I will discuss the fourth compound TI3D_0107_B13 later, since it has additional phenotypes.

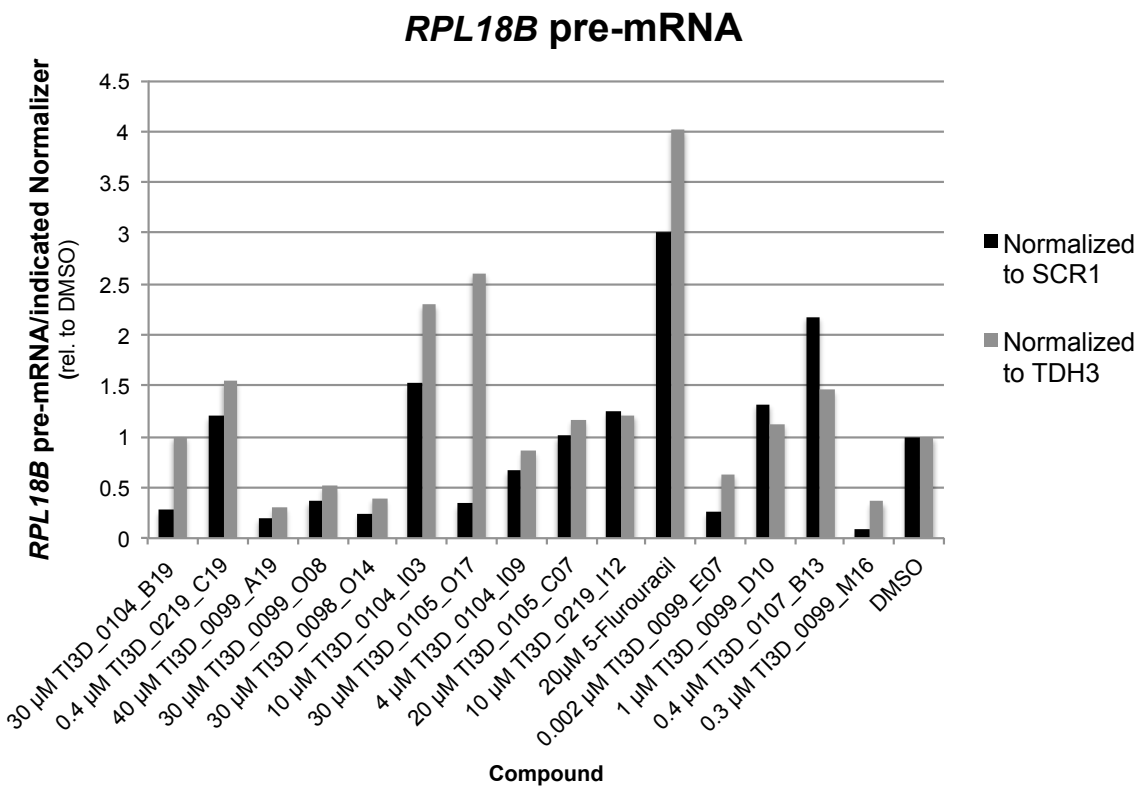


Figure 3.17: RT-qPCR analysis of the Rrp6p sensitive *RPL18B* pre-mRNA in the presence of small molecules

RT-qPCR analysis was performed with total RNA isolated from GM ABC-16 cells treated with the indicated concentration of compound for 6 hours. A standard curve technique was used to calculate the transcript levels and the *RPL18B* pre-mRNA is normalized to two different control transcripts, *SCR1* and *TDH3.* Normalized levels are charted relative to DMSO only treated cells. Since this experiment served as a quick screen, only one biological replicate was analyzed, therefore no error bars are shown.

As our reporter is sensitive to pre-mRNA splicing defects and some of the hit compounds give rise to a red-shifted/pre-mRNA defect-like phenograph, we also wanted to complete RT-qPCR to assay pre-mRNA splicing efficiency of endogenous introns. I queried three endogenous introns as well as our synthetic reporter intron. In looking at the results, the first finding that stands out is that there are obviously gene-specific effects (Figure 3.18). Of note, the efficiency of *DBP2* splicing has a wide range of defects without consistent patterns with the other introns. The two compounds with consistent splicing defects across all four introns are TI3D_0104_B19 and TI3D_0105_O17. As stated before, TI3D_0105_O17 is a well-characterized transcription inhibitor. Given the cotranscriptional nature of pre-mRNA splicing, I was not surprised to see a pre-mRNA splicing defect, as many transcription mutants display red-shifted reporter phenographs (Figure 2.3D). As for TI3D_0104_B19 there are no known links between its known mechanism of action and pre-mRNA splicing, making the further exploration of this compound a very interesting future research topic.

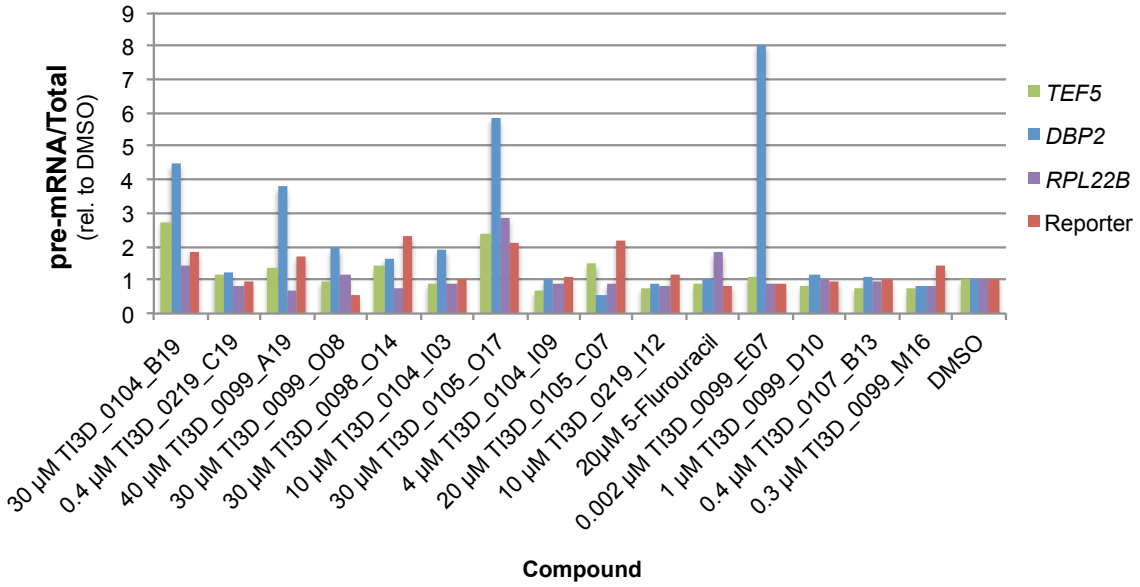
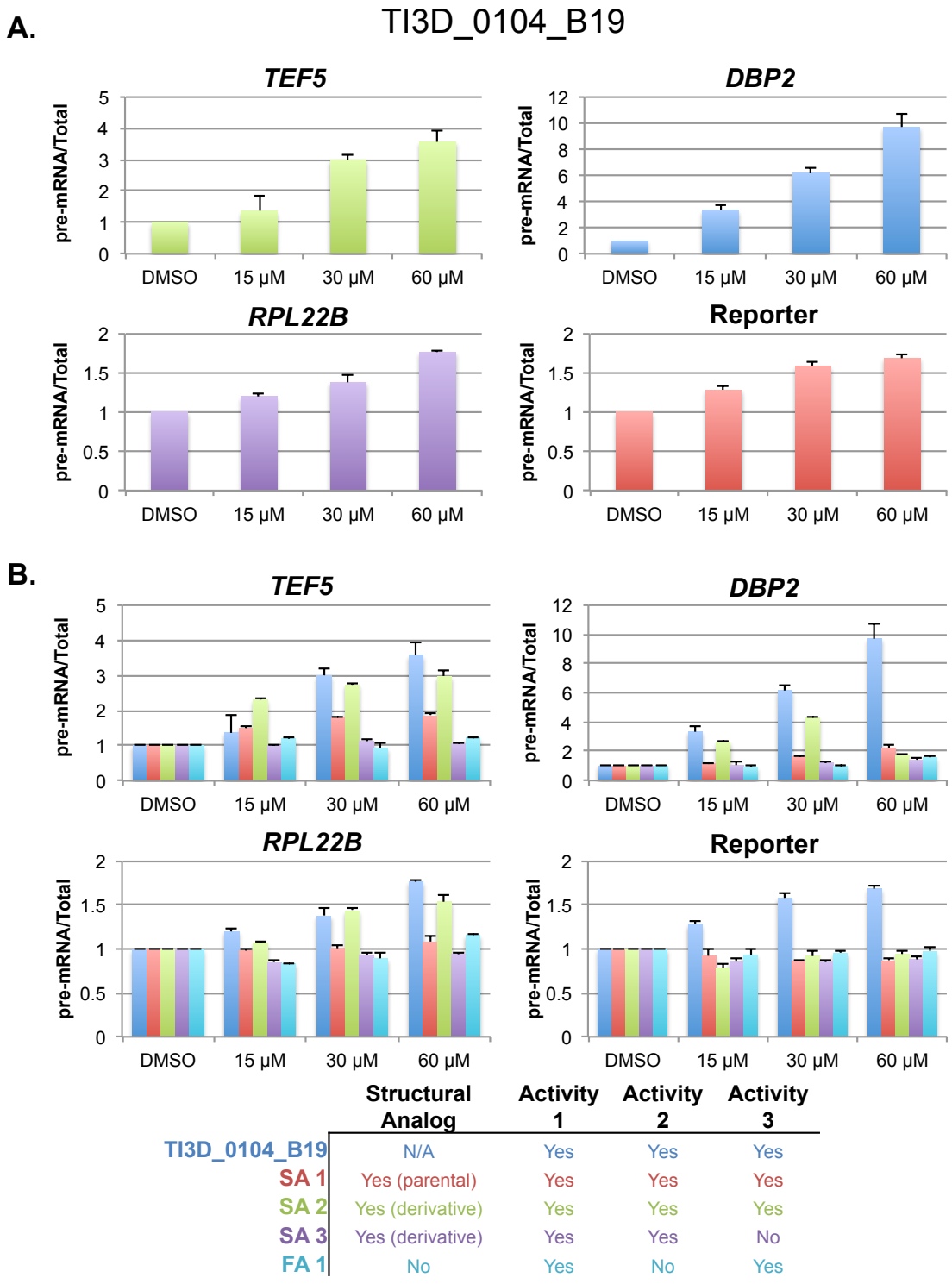


Figure 3.18: RT-qPCR survey for pre-mRNA splicing defects in small molecule treated cells

RT-qPCR analysis was performed with total RNA isolated from GM ABC-16 cells treated with the indicated concentration of compound for 6 hours. A standard curve technique was used to calculate the transcript levels. The pre-mRNA/total ratios are charted relative to DMSO only treated cells. As in Figure 3.17, only one biological replicate was analyzed.

To confirm the potential pre-mRNA splicing defect observed in TI3D_0104_B19 treated cells, I performed a similar RT-qPCR analysis, testing at three different concentrations (15, 30, and 60 μM) in biological triplicate (Figure 3.19A). I was very encouraged to detect dose-dependent pre-mRNA splicing defects for all four introns tested, with very strong 4- and 10-fold splicing defects for *TEF5* and *DBP2* introns respectively. Although the degrees of splicing defect were less for *RPL22B* and the reporter intron, the dose dependent response was still evident and the fold-changes were similar to splicing mutants.



	Structural Analog	Activity 1	Activity 2	Activity 3
TI3D_0104_B19	N/A	Yes	Yes	Yes
SA 1	Yes (parental)	Yes	Yes	Yes
SA 2	Yes (derivative)	Yes	Yes	Yes
SA 3	Yes (derivative)	Yes	Yes	No
FA 1	No	Yes	No	Yes

Figure 3.19: Identification of a novel dose-dependent pre-mRNA splicing defect in TI3D_0104_B19 treated cells

(Figure 3.19 continued)

(A) GM ABC-16 cells with copper inducible reporter were treated with three different concentrations of TI3D_0104_B19 and the reporter was induced for 6 hours. Total RNA was isolated and RT-qPCR was performed using primers in Table 3.3. All pre-mRNA/total ratios are taken relative to DMSO only and error bars representing one standard error. (B) Same as panel A, except for 3 different concentrations of five different compounds were tested. TI3D_0104_B19 was tested alongside three structural analogs (SA) and one functional analog (FA). The legend on the bottom describes the compounds in terms of structure and function relationship to TI3D_0104_B19, and color labels uniform throughout the panel.

Often, one can determine more about the mechanism of action of a compound by surveying compounds with a similar chemical structure or small molecules with overlapping activities. One of the four compounds I chose to test was the parental compound (SA 1) of TI3D_0104_B19, which shares all three known activities. I also chose to examine two additional structural analogs of TI3D_0104_B19, SA 2 and SA 3. SA 2 shares all three activities while SA 3 lacks activity 3. The last compound I tested, FA 1 (functional analog), does not have a similar structure, but displays activities 1 and 3. I treated drug-sensitive yeast with these four compounds at the same concentrations and performed RT-qPCR analysis for the same four introns as in Figure 3.19A. FA 1 and SA 3 did not have an appreciable affect on the splicing efficiency of any of the introns tested. SA 1, the parental compound of TI3D_0104_B19, had a very slight splicing defect at 30 and 60 μM for *TEF5* and *DBP2* introns, but not for the *RPL22B* and reporter introns. The one other compound that did have an appreciable negative affect on the splicing efficiency of all three endogenous introns was SA 2, although the overall effects were muted compared to TI3D_0104_B19. Altogether, the general splicing inhibition seen with TI3D_0104_B19 treatment is specific to this chemical structure and potentially all three known activities. Future efforts to study the splicing inhibition activity of this compound can be found in the Discussion (Section 3.4).

3.4 DISCUSSION

In this chapter I describe two complementary approaches to discover small molecules that modulate yeast gene expression using our versatile gene expression reporter. Instead of a standard *in vitro* high-throughput screen using purified target protein, both of our approaches allowed us to survey the effect on many different gene expression processes simultaneously in the context of a biological system (cell). Both the cyclic peptide and small molecule screens yielded putative splicing inhibitors, as well other molecules of interest than can be explored further. One general observation in both screens is that cyclic peptides and small molecules can elicit more heterogeneous effects (Figures 3.1B and 3.14). In other words, the reporter phenograph populations are more disperse due to increase variability in reporter signal. This potentially shows the result of partial inhibition in contrast with a clean deletion of a component of gene expression machinery.

In my cyclic peptide screen I discovered that it is prone to a relatively high false-positive rate. One contributing factor to this was a 96-well plate analysis artifact, meaning that I observed significant shifts when analyzing potential hits in 96-well plates (high-throughput mode), but when analyzing carefully from culture tubes, this phenotype was absent. Although it is nearly impossible to control for 96 different growth rates in a single plate, we have overcome this by analyzing plates at different time points and measuring the average absorbance of the plate. One could also use a plate reader to quickly analyze the growth stage of a

given well and select promising wells for flow cytometry analysis at different time points, analyzing the wells that are at the ideal optical density (A_{600}). As with any yeast genetic screen, false positives are inevitable, but to have a system to either eliminate them or quickly rule them out is of great importance.

In surveying the expression and processing of the SICLOPPS-encoded cyclic peptides, we are relying on the chitin-binding-domain (CBD) fused to the C-terminal portion of the split intein. Although this is useful, it does not truly measure the levels of resulting cyclic peptide, just the precursor and byproduct (Illustration 3.1C). Having a way of detecting the cyclic peptide, such as the incorporation of an unnatural amino acid, would help determine *in vivo* levels of cyclic peptide. This being said, the SICLOPPS system is a very attractive tool. One of the features that set SICLOPPS apart in small molecule screening is the ability to rapidly study the structure-activity relationship of the hit cyclic peptides by straightforward site-directed mutagenesis, which is much simpler than the chemical synthesis of modified compounds.

We are disappointed that the cellular target for the red-shifted hit R6_E04 is still unknown. Although we were able to validate a small but reproducible pre-mRNA splicing defect by RT-qPCR analysis of endogenous introns (Figure 3.9), we do not currently have any data that points to a specific target. With the attempted pull-down experiments using the biotinylated cyclo-CSEKGAGL (R6_E04 V4K), there would need to be a relatively strong interaction between the cyclic peptide and its target for us to identify. Further optimization could be useful

to further clear the extract as well as adjust the stringency of the washes, etc. It is interesting to note that we observed increased activity in the V4K mutant compared to the parental R6_E04 (Figure 3.5C), which could indicate that the lysine residue plays a role in its activity. Attaching a biotin moiety would most likely be detrimental to its activity if this were the case. On the other hand, the V4K clone did not have any splicing defect when looking at endogenous introns (Figure 3.9), hinting towards a possibility that R6_E04's activity is specific to our reporter.

As a complementary approach to the biochemical target identification, I also performed a FACS screen to identify high copy suppressors using a plasmid based genomic tiling library, but my efforts were unsuccessful. An additional fairly straightforward experiment would be to repeat the pull-down experiment and probe for probable RNAs that the R6_E04 might interact with, such as snRNAs or intron-containing pre-mRNAs. Furthermore the biotinylated cyclo-CSEKGAGL might be useful to probe for treating extract with the cyclic peptide and spliceosome purification. Of course a strong interaction would be required for this experiment to work.

In terms of the high-throughput small molecule screen, I believe the pilot efforts made were encouraging, warranting additional screening. The libraries that I screened were highly populated with existing and failed drugs. As the discovery of novel drugs slows, there is a growing trend of drug repurposing or repositioning (Corbett et al. 2012; Gupta et al. 2013; Jin and Wong 2014; Cragg

et al. 2014). The idea is to find novel uses for existing drugs that might have been overlooked before, or combining drugs for synergistic effects. Our reporter assay, with its wide yet specific sensitivity, is ideal for this type of work to identify currently unknown activities of existing drugs.

In moving forward in screening efforts, it would be wise to establish additional orthogonal assays to validate and further test potential hit compounds, which will also get closer to identifying a cellular target. When using flow cytometry, it is also possible to multiplex assays and screen against multiple cell types or reporters simultaneously. This can be done by labeling cells with different markers or combinations of markers and using them to gate the different cells types, in which you can then assay your signal of interest (Chen et al. 2012). In terms of target identification, I believe our binning and clustering analysis with gene expression deletion mutants will continue to work well in narrowing down the number of pathways (Figures 3.8 and 3.16). An additional tool in yeast drug discovery is the use of chemical genetic interactions (Giaever et al. 2004; Parsons et al. 2006; Hillenmeyer et al. 2008), in which one screens yeast deletion collections or overexpression libraries for genes that lead to differential growth in the presence of a given compound.

Out of over three thousand small molecules, our screen identified TI3D_0104_B19, a novel pre-mRNA splicing inhibitor, as resulting in a red-shifted phenograph. Not only did the compound treated sample cluster with pre-mRNA splicing mutants (Figure 3.16A), I also discovered dose-dependent

splicing impairment of endogenous introns (Figure 3.19). Given the nature of this small molecule and its known functions, it is possible that this molecule acts upstream of pre-mRNA splicing and might have indirect effects. There is one published work that links TI3D_0104_B19 to alternative splicing in mammalian cells, but the molecular details are unclear. Further experimentation is needed to determine if this molecule directly inhibits pre-mRNA splicing, and if so, which spliceosome component. The availability of a radioisotope-labeled version of this compound, as well as its fluorescence properties, can aid in these efforts. It would also be interesting to determine the nature behind the gene-specific effects of some of the structural derivatives tested. Optimization of this compound in terms of pre-mRNA splicing will most likely be needed since most promising small molecules have activity in the low- to sub-micromolar range, compared to the mid-micromolar range observed in yeast. It would not be surprising though to see increased efficacy in mammalian cells if permeability and stability are increased. The fact that pre-mRNA splicing is becoming a more sought after target for chemotherapeutics is encouraging as well (Dehm 2013; Effenberger et al. 2014).

One gene expression target that our reporter assay is particularly sensitive to is the TREX-2 complex, which functions in linking mRNA and export (Fischer et al. 2002; Rodriguez-Navarro et al. 2004). As seen in Figure 2.4A, deletion of the nonessential components of the TREX-2 complex gives rise to a strong a reproducible green-shift. As discussed earlier (Section 3.1.5) the human protein

GANP of the TREX-2 complex is strongly implicated in cancer and has been shown to be a valuable chemotherapeutic target (Fujimura et al. 2005; Kageshita et al. 2006; Wickramasinghe et al. 2010; Phimsen et al. 2012). Very recently TREX-2 components have been shown to be involved in DNA damage, genome instability and R-loop formation via BRCA2 (Bhatia et al. 2014). The medical relevance of the TREX-2 complex and the unique signal in TREX-2 yeast mutants further highlight the utility of our reporter as a drug-screening tool.

I am confident that the cyclic peptide library (including different sizes) is still a valuable tool for further exploration, as I have not come close to exploring the entire CX7 library. Given the results so far, I believe that these libraries contain biological interesting cyclic peptides that can be identified with our reporter. In terms of small molecules, our pilot screens of over 3000 compounds is barely scratching the surface. With the installation of a truly high-throughput flow cytometry system on campus, our optimized reporter system is ready to survey tens or hundreds of thousands of molecules. With these two screens, I have further demonstrated the broad utility of our reporter in surveying the effect of small molecules on a wide range of gene expression process simultaneously.

Chapter 4: Rapid identification of loss-of-function mutants of Swm2p

4.1 Background

In eukaryotic cells, the accurate expression of mRNA requires the coordinated efforts of multiple cellular machines to generate, process, and translate the desired protein (Komili and Silver 2008). One such machine, the spliceosome, is a large and dynamic complex comprised of five small nuclear ribonucleoproteins (snRNPs), each consisting of a small nuclear RNA (snRNA) and associated proteins. These snRNPs function in pre-mRNA splicing to remove noncoding introns to generate translatable mRNAs (Nilsen 2003). Although there are differences between the snRNPs that make up the spliceosome, their biogenesis is a tightly orchestrated process; most often including snRNA processing and modifications, as well as nucleocytoplasmic shuttling and packaging (Kiss 2004; Patel and Bellini 2008; Fischer et al. 2011). The timely and accurate biogenesis of snRNPs, as well their dynamic interactions with the pre-mRNA substrate are essential for an efficient pre-mRNA splicing and gene expression program.

4.1.1 Modifications of snRNAs

The mechanisms and functions of spliceosomal snRNA modifications have been extensively studied and reviewed (Karijolich and Yu 2010), yet there is still much to learn about their functions and biological implications. All five

snRNAs are modified posttranscriptionally, including 2'-*O*-methylation, pseudouridylation and 5' cap trimethylation (Reddy and Busch 1988; Grosjean and Benne 1998; Massenet et al. 1999; Grosjean 2005). The internal modifications to both the backbone (2'-*O*-methylation) and bases (pseudouridylation) are highly conserved and are mainly located in functionally important regions (Szkukalek et al. 1995; Gu et al. 1996; Yu et al. 1998; Patton and Padgett 2005; Simoes-Barbosa et al. 2008). Although the majority of snRNA modifications are catalyzed via RNA-dependent mechanisms (Box H/ACA or Box C/D snoRNAs and their associated proteins), some are performed by enzymes in a RNA-independent manner.

4.1.2 Trimethylguanosine synthase

In terms of 5' cap trimethylation, U6 is the only snRNA (RNA Pol III transcript) that is not trimethylated. After synthesis by RNA polymerase II, the m^7G cap of the U1, U2, U4, and U5 snRNAs, as well as some small nucleolar RNAs (snoRNAs; Mouaikel et al. 2002) and the telomerase RNA (Jády et al. 2004; Franke et al. 2008), are converted to the 2,2,7-trimethylguanosine (TMG) cap structure. This reaction is catalyzed by trimethylguanosine synthase (Tgs1) via two successive methylations using the methyl donor *S*-adenosylmethionine (Zhu et al. 2001; Mouaikel et al. 2002; Komonyi et al. 2005; Hausmann and Shuman 2005; Hausmann et al. 2007, 2008; Benarroch et al. 2010). Although the biogenesis of snRNAs and the localization of specific events have been well

characterized in human cells, there is still much that is unknown about snRNA biogenesis in the budding yeast.

Even though Tgs1 is essential in higher eukaryotic organisms (Komonyi et al. 2005), Tgs1p is not required for viability in *Saccharomyces cerevisiae,* yet yeast lacking *TGS1* exhibit a cold-sensitive growth phenotype (Mouaikel et al. 2002). The deletion of *TGS1* results in splicing and ribosomal RNA (rRNA) processing defects, as well as decreased nucleolar organization (Mouaikel et al. 2002; Colau et al. 2004; Boon and Kos 2010). The nonessential nature of Tgs1p in budding yeast argues for functional redundancy or the lack of a strict need for the TMG cap in snRNA biogenesis. Recently it has been shown that Tgs1p is essential for meiosis in yeast, being responsible for the splicing of the *SAE3* and *PCH2* meiotic transcripts (Qiu et al. 2011).

In higher eukaryotes, U1, U2, U4, and U5 snRNAs are transported to the cytoplasm after acquiring a m^7G-cap (Hernandez 2001; Cougot et al. 2004). The Sm-proteins then assemble around the Sm-site (Plessel et al. 1994; Kambach et al. 1999b, 1999a), and are a requirement for TMG cap formation (Mattaj 1986; Plessel et al. 1994). The trimethylated cap and Sm-core then act as import signals for the snRNPs to return to the nucleus (Mattaj and De Robertis 1985; Hamm et al. 1990; Fischer et al. 1993; Palacios et al. 1997; Huber et al. 1998; Narayanan et al. 2002, 2004). In contrast, there is no evidence of a nucleocytoplasmic transport cycle for snRNAs in yeast or the export and import adaptors required for snRNA/snRNP shuttling. Additionally Tgs1p is localized in

the yeast nucleolus (Mouaikel et al. 2002), which is most likely where cap trimethylation takes place.

One open question regarding snRNA trimethylation in yeast concerns Tgs1p target specificity. There is some evidence supporting Tgs1p targeting to snRNAs via protein-protein interactions with snRNP components (Mattaj 1986; Plessel et al. 1994; Mouaikel et al. 2002). More recently, Tgs1p has been found to interact physically and genetically with factors involved in splicing and RNA processing (Hausmann et al. 2008; Wilmes et al. 2008; Chang et al. 2010; Boon and Kos 2010). More specifically, large-scale two-hybrid screens and directed studies have identified the interacting partner Swm2p (synthetic with Mud2; Uetz et al. 2000; Yu et al. 2008; Chang et al. 2010; Boon and Kos 2010). The nonessential *SWM2/YNR004W* is an interesting gene that codes for a fairly poorly characterized small protein (17 kDa; 146 amino acids) that contains no recognizable domains (Volckaert et al. 2003) and has only three plausible fungal homologs (Wilmes et al. 2008). Of note, the deletion of *TGS1* has a similar genetic interaction profile to *SWM2* null cells (Wilmes et al. 2008), however *SWM2* and *TGS1* do not appear to genetically interact (Chang et al. 2010; Boon and Kos 2010). More recently it has been shown that Swmp2 interacts with the N-terminus of Tgs1p and that its deletion results in a loss of U3 and snRNA trimethylation, with no effect on a number of snoRNAs (Boon and Kos 2010). Additionally, those authors discovered that Tgs1p was required for nucleolar localization of Swm2p, but not vice versa. These findings lead to a hypothesis

that Swm2p may function as a specificity factor for Tgs1p by targeting it to the Pol II generated snRNAs.

4.1.3 Experimental approach and rationale for *SWM2* mutagenesis screen

In our previously published screen of the yeast deletion collection with a gene expression reporter, we recognized that the deletion strain lacking *SWM2* had an increase in the pre-mRNA/mRNA ratio, indicating a potential splicing defect (Figure 2.3A; Sorenson and Stevens 2014). This sparked our interest since the majority of yeast splicing factors have been identified and scrutinized in detail. During the course of our initial work, two research articles shed light on the potential function of Swm2p (Chang et al. 2010; Boon and Kos 2010), but the molecular details behind this potential function were still unclear. With little idea of any functional domains or regions of interest within the *SWM2* gene, we employed a standard mutagenesis approach coupled with our gene expression reporter to quickly and efficiently screen tens of thousands of mutants via fluorescence-activated cell sorting to identify functionally interesting mutants. Our efforts resulted in three *swm2* mutants that had reproducible yet gene specific effects on pre-mRNA splicing and reduced interaction with Tgs1p.

4.2 MATERIALS AND METHODS

4.2.1 Construction and culturing of yeast strains

Standard yeast methods were used (Sherman 1991) and yeast were grown at 31°C unless otherwise noted. Yeast strains grown in 96-well plates for screening purposes were cultured in flat bottom 96-well plates (Corning) with 150 µL of broth on a microplate shaker (VWR) at 2000 rpm. Yeast strains used in this chapter can be found in Table 4.1.

To generate the genomic *SWM2*-TAP, strains were generated by homologous recombination. First, *TRP1* was deleted from the wild-type BY4741 strain (SS4004) using the *MPY-URA3* replacement method (Schneider et al. 1996) to generate the more versatile wild-type strain SS4019. Briefly, an *MPY-URA3* PCR fragment generated to target *TRP1* (Plasmid MS112; Oligos MS216/217) was transformed via standard lithium acetate transformation (Geitz and Scheistel 1995). Transformants were selected in SD plates lacking uracil. Verified strains were then patched on YPD overnight, then streaked onto synthetic complete plates with 5-FOA to select for cells in which an *MPY* recombination event occurred, looping out the incorporated *URA3* at the *TRP1* locus. The resulting strain SS4019 was transformed with TAP-tagging/truncation PCR products using template MS138 and oligos MS164-171 to generate C-terminal truncation Swm2p-TAP strains SS4038-4044. The Swm2p-TAP clone from the TAP-tag collection was used for the full-length strain SS4037 (Howson et al. 2005). The *mud2*Δ*/swm2*Δ strains (SS4147-4150) were obtained from

Beate Schwer (Chang et al. 2010). Plasmids and DNA oligonucleotide sequences used in strain construction can be found in Tables 4.2 and 4.3 respectively. The *swm2*Δ Tgs1p-myc13 strain (SS4153) was made by homologous recombination of a PCR product using oligos MS656/657 and plasmid MS419.

Table 4.1 Yeast strains used in Chapter 4

Strain (SS)	Mating type	Genotype
4004	a	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* (BY4741)
4019	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0* (BY4741 *trp1Δ*)
4037	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 SWM2*-TAP-*HIS3MX6*
4038	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC15-TAP-*TRP1*
4039	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC30-TAP-*TRP1*
4040	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC45-TAP-*TRP1*
4041	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC60-TAP-*TRP1*
4042	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC75-TAP-*TRP1*
4043	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC90-TAP-*TRP1*
4044	a	*his3Δ1 leu2Δ0 met15Δ0 trp1Δ ura3Δ0 swm2*-ΔC105-TAP-*TRP1*
4052	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX* + pRS316 *TDH3*-Reporter (MS127)
4053	α	his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 *swm2Δ::*KanMX + pRS316 *TDH3*-Reporter (MS127) + pRS415 *SWM2*-TAP (MS161)
4059	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2*Δ*::KanMX*
4060	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 tgs1*Δ*::KanMX*
4147	α	*his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*
4148	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 mud2*Δ*::natR*
4149	a	*his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 swm2*Δ*::kanR*

(Table 4.1 continued)

4150	α	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mud2Δ::natR swm2Δ::kanR*
4153	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX TGS1-13myc::HIS3MX6*
4154	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX TGS1-13myc::HIS3MX6* + pRS415
4155	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX TGS1-13myc::HIS3MX6* + pRS415 *SWM2*-TAP
4156	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX TGS1-13myc::HIS3MX6* + pRS415 *swm2-P4D9*-TAP
4157	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX TGS1-13myc::HIS3MX6* + pRS415 *swm2-P4E9*-TAP
4158	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX TGS1-13myc::HIS3MX6* + pRS415 *swm2-P4H7*-TAP
4159	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX* + pRS415
4160	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX* + pRS415 *SWM2*-TAP
4161	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX* + pRS415 *swm2-P4D9*-TAP
4162	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX* + pRS415 *swm2-P4E9*-TAP
4163	α	*his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 swm2Δ::KanMX* + pRS415 *swm2-P4H7*-TAP
4164	α	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mud2Δ::natR swm2Δ::kanR* + pRS415
4165	α	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mud2Δ::natR swm2Δ::kanR* + pRS415 *SWM2*-TAP
4166	α	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mud2Δ::natR swm2Δ::kanR* + pRS415 *swm2-P4D9*-TAP
4167	α	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mud2Δ::natR swm2Δ::kanR* + pRS415 *swm2-P4E9*-TAP
4168	α	*his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mud2Δ::natR swm2Δ::kanR* + pRS415 *swm2-P4H7*-TAP

Table 4.2 Plasmids used in Chapter 4

MS#	Common Name	Purpose	Source
102	pRS415	Yeast Expression Vector	(Sikorski and Hieter 1989)
112	p*MPY-URA3* (p*MPY*-ZAP)	Gene knockout via removable *URA3*	(Schneider et al. 1996)
127	pRS316 *TDH3*-Reporter	*TDH3* driven reporter	This work.
138	pBS1479 (TAP-*TRP1*)	TAP tagging via *TRP1*	(Puig et al. 2001)
159	TA *SWM2*-TAP	Cloning *SWM2*-TAP	This work
161	pRS415 *SWM2*-TAP	Parental *SWM2*-TAP plasmid	This work
395	pRS415 *swm2-P4D9*-TAP	*swm2* mutant from screen	This work.
396	pRS415 *swm2-P4E9*-TAP	*swm2* mutant from screen	This work.
397	pRS415 *swm2-P4H7*-TAP	*swm2* mutant from screen	This work.
419	pFA6a 13myc-*His3*MX6	C-terminal myc tagging vector	(Longtine et al. 1998)

Table 4.3 DNA oligonucleotides used in Chapter 4

#	Sequence (5' to 3')	Purpose
129	CCTGTTATCCCTAGCGGATCTG	GFP/TAP-*ADH1*term PCR3
154	TAAAACCTATAGAAACTCCTTC	*SWM2* PCR5
155	TTAATGATGAACTCGAAAGAGC	*SWM2* PCR3
164	GGAATTTGACGAATTTGGTAAACTCATAAGTAATCCATTAGAGTCCATGGAAAAGAGAAG	*SWM2*del15 TAPA
165	TGAGATTGATGTCTCATATGTTAAGAATTTGAATGTAGTCAGGTCCATGGAAAAGAGAAG	*SWM2*del30 TAPA
166	TAAAAAGTTGAAGGATCCTCCGCCTGCTGATAATATCATTAAGTCCATGGAAAAGAGAAG	*SWM2*del45 TAPA
167	AAGAAGATGCCTTAACCTGTGGTATGTTTTGAAGGGCAAGGAATCCATGGAAAAGAGAAG	*SWM2*del60 TAPA

(Table 4.3 continued)

168	TAAATGGAACGAAGAAGGTGCGAAGACAGTCAATCCGGAAATTTCCATGGAAAAGAGAAG	*SWM2*del75 TAPA
169	ACATTTAAAGTCTAGAGCTCTAAAAATTTGTCGTTCTAATTTCTCCATGGAAAAGAGAAG	*SWM2*del90 TAPA
170	CTCCGCAGTATTGGAGCCATATTTTCAGAATATAGCAAGAAATTCCATGGAAAAGAGAAG	*SWM2*del105 TAPA
171	AAAGCTAAAAGGTAGGCTCCGTGCATATTAGATGAGTGAACACTACGACTCACTATAGGG	*SWM2* TAPB
178	AACTCCTTCATATAAAATGTACTAG	*SWM2* mut PCR fwd
179	CCATGGATTAATTAACCCGGGGATC	*SWM2* mut PCR rev
216	TATTGAGCACGTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTAGGGAACAAAAGCTGG	*TGS1-MPY URA3* UP
217	TGCAGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAACCTATAGGGCGAATTGG	*TGS1-MPY URA3* DOWN
282	CCATTGTCACTCTAGATGTCAAGCC	*TEF5* total FWD qPCR
283	GATACCGAAACCAATTGGGATAAAT	*TEF5* total REV qPCR
299	NNNNNNNNG	NonamerG for RT
300	NNNNNNNNA	NonamerA for RT
301	NNNNNNNNT	NonamerT for RT
302	NNNNNNNNC	NonamerC for RT
313	GATAGCACAGAGCAGAGTATCATTA	*TEF5*intFWDjp qPCR
314	CTGGAGAATTCTGGGTAAGCAGATT	*TEF5*intREVjp qPCR
584	ACAGAATAACGAACCAAATTACTAACAGT	*DBP2*intFWDms qPCR
585	CCTGGCATATCGTAGTTGATAACG	*DBP2*intREVms qPCR
586	CTTCACCGAACAAAACAAAGGTT	*DBP2*exonFWDms qPCR
587	TCGGGAGGAATATTTTGATTAGCT	*DBP2*exonREVms qPCR
606	GGGAAAATCAGCGTATGTAAAGC	*SRB2* pre-mRNA FWD qPCR

(Table 4.3 continued)

607	GAGTGGCTCTTTCCACGAATATAA	*SRB2* pre-mRNA REV qPCR
608	CATCGAAGGACACCTAGCTGAA	*SRB2* total FWD qPCR
609	TCCGGCCCCAACGAG	*SRB2* total REV qPCR
624	CAAATAGGGTGGGACCAACA	*RPS21B* pre-mRNA FWD qPCR
625	TCAGGTAACACTTGTGCCAAT	*RPS21B* pre-mRNA REV qPCR
626	ATCCAGAGGTGAATCCGATG	*RPS21B* total FWD qPCR
627	AACGACTTCCCCTCTTCTTTTT	*RPS21B* total REV qPCR
640	GACCCAGGAAAGCATCTTAC	*snR4* Northern
645	GATAGTATTAACCACTACTG	*snR37* Northern
646	GAATGGAAACGTCAGCAAACA	U1 Northern (*snR19*)
647	AAAAGAACAGATACTACACTTGA	U2 Northern (*LSR1*)
648	ACCATGAGGAGACGGTCTGG	U4 Northern (*snR14*)
649	ACACCCGGAATGGTTCTGGTA	U5 Northern (*snR7-S/L*)
650	AACGAAATAAATCTCTTTGTAAAAC	U6 Northern (*snR6*)
656	GACCAAAGATAATACTGTAGACATT TACGACGTAAATGGTCGGATCCCC GGGTTAATTAA	*TGS1* F2
657	CATTGCATTTTCTATCATACGATTC AACTTGAACAAGTGTGAATTCGAG CTCGTTTAAAC	*TGS1* R1

4.2.2 *SWM2* cloning and plasmid generation

The *SWM2*-TAP open reading frame, along with its native promoter, was amplified using genomic DNA from the corresponding tap collection strain (Howson et al. 2005) and primers MS129/154, cloned and sequenced verified in a cloning vector (pGEM-T Easy; Promega; MS159), then cloned into pRS415 using NotI, resulting in pRS415 *SWM2*-TAP (MS161). Site-directed mutagenesis was used to introduce individual point mutants. Plasmids and DNA oligonucleotide sequences can be found in Tables 4.2 and 4.3 respectively.

4.2.3 Generation of *swm2* mutant pool

To generate random *swm2* mutants, 1 ng pRS415 *SWM2*-TAP (MS161) was used with 1 μM primers MS178/179 for mutagenic PCR (Beckman et al. 1985; Cadwell and Joyce 1992) with Taq Polymerase (NEB; 1 unit), 300 μM $MnCl_2$, 1 mM dNTPs, and 1X NEB Thermo Pol Buffer. After 18 cycles the PCR product was gel purified and transformed into *swm2Δ* yeast (with pRS316 *GPD*-Reporter) alongside restriction enzyme (NdeI and NsiI) treated/gel purified MS161 for gap-repair (Gietz and Schiestl 2007; Lundblad and Struhl 2008). During optimization, random mutant clones were analyzed to verify a mutation rate of about 1.5 nucleotide substitutions per *SWM2* open reading frame. Over 20,000 transformants were obtained on synthetic media lacking leucine and uracil, pooled, and then used to inoculate a culture to screen via FACS (remaining mutant pool was cryopreserved).

4.2.4 Mutagenesis screen using reporter and FACS

For fluorescence-activated cell sorting, WT cells (*swm2Δ* + pRS415 *SWM2*-TAP and pRS316 *GPD*-Reporter) were used to determine gating boundaries. Immediately before sorting, the mutant pool (grown to OD (A_{600}) 0.5) was resuspended in sterile 1X PBS. A FACSAria (Becton Dickinson) was used to sort 37K red-shifted cells into synthetic media lacking leucine and uracil. 2000+ mutants were plated for individual colonies. Once grown, single clones were transferred to 96-well plates for storage and subsequent high-throughput flow

cytometry analysis to identify mutant clones that were reproducibly and significantly red-shifted. Expression of Swm2p-TAP was verified by western blot analysis. *SWM2* plasmids from clones of interest were isolated using the Zymoprep Yeast Plasmid MiniprepII kit (Zymo) after shuffling out the *URA3* reporter with 5-fluoroorotic acid. Isolated plasmids were transformed into *E. coli*, sequence verified, and retransformed into *swm2Δ* yeast with the pRS316 *GPD*-Reporter to verify effect.

4.2.5 Western blot analysis

Protein samples were separated on PAGE gels and transferred to nitrocellulose. Membranes were blocked with BLOTTO (PBST + 5% SDS) and probed with 1:5000 peroxidase anti-peroxidase antibody (Rockland) for the TAP tag. Probing for the 13myc tag was performed with 1:2500 anti-c-myc purified from Myc 1-9E10.2 (ATCC CRL-1729) followed by 1:5000 goat anti-mouse with HRP (Rockland 610-1302). Washes were preformed with PBST and visualized using X-ray film using chemiluminescence (Perkin Elmer). For the TAP coimmunoprecipitations chemifluorescence reagent was used (GE, ECL Plus) and signals were detected using a GE Typhoon scanner. Quantification of TAP and myc signals was performed using ImageQuant (GE Healthcare) with background subtraction.

4.2.6 Trimethylguanosine immunoprecipitation

Immunoprecipitation for 2,2,7-trimethylguanosine (TMG) capped RNA was performed using pulverized cell material in 1 mL/g AGK (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 8% glycerol). About 1 gram of extract was thawed on ice (added 1 µM BME, 0.4 mM PMSF, 5 µM leupeptin, and 3 µM pepstatin) and cleared by centrifugation (20 minutes, 14,000 x *g*, 4°C) in a microfuge tube. 100 µL cleared extract was transferred to a microfuge tube with 5 µL α-TMG antibody (K121, Santa Cruz) and rotated for 1 hour at 4°C. The extract/antibody mixture was transferred to a tube containing 25 µL pre-washed Protein G agarose beads (50 µL slurry, Invitrogen, washed with AGK) and rotated for 2 hours at 4°C. Samples were centrifuged (3 min., 3000 x *g*, at 4°C) and the supernatant was removed (saving 10 µL). Beads were quickly washed two times with 1 mL AGK, and washed a third time for 10 minutes with rotation at 4°C. After beads were centrifuged and wash was removed, 100 µL AGK was added to the beads (90 µL AGK was added to 10 µL samples of cleared extract and supernatant). Two phenol chloroform extractions were preformed and the RNA was precipitated overnight at -20°C with sodium acetate and ethanol (glycogen added to IP samples). RNA was washed twice with cold 70% ethanol, dried, and resuspended in DEPC ddH_2O (15 µL for IP and 30 µL for total and supernatant). 3 µL RNA was used for northern blot analysis.

4.2.7 Northern blot analysis

Equal volumes of RNA and 8M Urea RNA loading buffer (8M urea, 20 mM EDTA pH 8.0, 2 mM Tris pH 7.5, 1.68 mg/mL xylene cyanol, 0.56 mg/mL bromophenol blue) were heated at 65°C, then electrophoresed in a 7% 19:1 polyacrylamide gel (8M urea 1X TBE) for 1 hour and 45 minutes at 25 mA. RNA was then transferred to a nylon membrane (Whatman) via downward capillary transfer (0.01 N NaOH and 3 M NaCl). After crosslinking by UV light (254 nm for 1 minute) the membrane was probed with 500,000 counts ^{32}P-*γ* end-labeled DNA (boiled immediately prior to addition; see Table 4.3 for DNA oligos) in Church Buffer (500 mM Na_2HPO_4 pH 7.0, 7% SDS, 1 mM ETDA pH 8.0) overnight. The blot was then exposed to a phosphorimaging screen and developed in a phosphorimager. Quantification of specific signals was performed using ImageQuant (GE Healthcare) with background subtraction. IP efficiency was calculated by normalizing IP signals to the total signal for each RNA species and then calculating efficiency relative to wild-type.

4.2.8 TAP coimmunoprecipitation

The procedure for TAP coimmunoprecipitation was based on the original work describing the TAP purification method (Puig et al. 2001). Briefly, 500 mL to 1 L of cells were grown to OD (A_{600}) 0.8 to 1.0 and frozen in 1 mL/g AGK (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 8% glycerol). After cryogenically pulverizing the material, about 1.25 grams of extract was thawed on ice in a

microfuge tube (added to 1 μM DTT, 0.4 mM PMSF, 5 μM leupeptin, and 3 μM pepstatin). Proteins were extracted by rotation at 4°C for 30 minutes. The extract was then cleared by centrifugation (20 minutes, 14,000 x *g*, 4°C). 500 μL cleared extract (normalized by OD (A_{595}) via Bradford assay) was transferred to a tube with 25 μL pre-washed IgG sepharose resin. 50 μL cleared extract was saved for "total" sample. The extract and resin were rotated for 1-2 hours at 4°C. 50 μL from the supernatant was saved for the "supernatant" sample after centrifuging 3000 x *g* for 3 minutes. Resin was then washed four times with 1 mL IPP150 (10 mM Tris, pH 8.0, 150 mM NaCl, 0.1% NP-40, 1.5 mM $MgCl_2$, 8% glycerol, 1 mM DTT) at 3000 x *g*. The bound proteins were eluted by boiling in 2X LDS + 200 μL DTT and saved for western blot analysis. For the total and supernatant samples, 50 μL of AGK was added and a phenol chloroform extraction was done, precipitating the organic phase overnight at -20°C with ice-cold acetone. Precipitated proteins were washed twice with 80% ice-cold EtOH, dried on the bench, then resuspended in 100 μL 2X LDS + 200 μL DTT. 30 μL of total and supernatant, and 1 μL IP (+ 19 μL loading buffer) were used for western blot analysis.

4.3 RESULTS

4.3.1 Rapid mutagenic screen of *SWM2* using our gene expression reporter and FACS

Our experimentation began by performing a mutagenic screen of *SWM2* utilizing standard error-prone PCR and gap-repair methods within a yeast strain lacking *SWM2* and harboring our gene expression reporter. Our efforts resulted in over 20,000 transformants/clones that were pooled and subsequently analyzed. When pre-mRNA splicing is impaired, such as in the case of *swm2Δ* (Figure 4.1A), there is an increase in the mCherry/GFP ratio within cells, shifting the population of cells toward the y-axis (mCherry) of our standard flow cytometry phenographs (GFP on x-axis). We used the gene expression reporter phenographs of wild-type cells and *swm2Δ* cells to determine the gating boundaries for the mutant pool (Figure 4.1A). Our goal was to sort mutant clones that were red-shifted, which would indicate impairment of pre-mRNA splicing or a related process (Sorenson and Stevens 2014). It was not surprising that very few cells within the mutant pool fell within this gate (right panel of Figure 4.1A), since most mutations would not likely negatively affect the function of Swm2p. This being said, we gated 0.4% of the population that fell into this gate and obtained more than 37,000 clones of interest within 15 minutes, most of which were cryopreserved for future analysis. In all over 10 million yeast were analyzed

which excessively, yet effortlessly, covered our original mutant pool of 20,000 clones 500 times over.

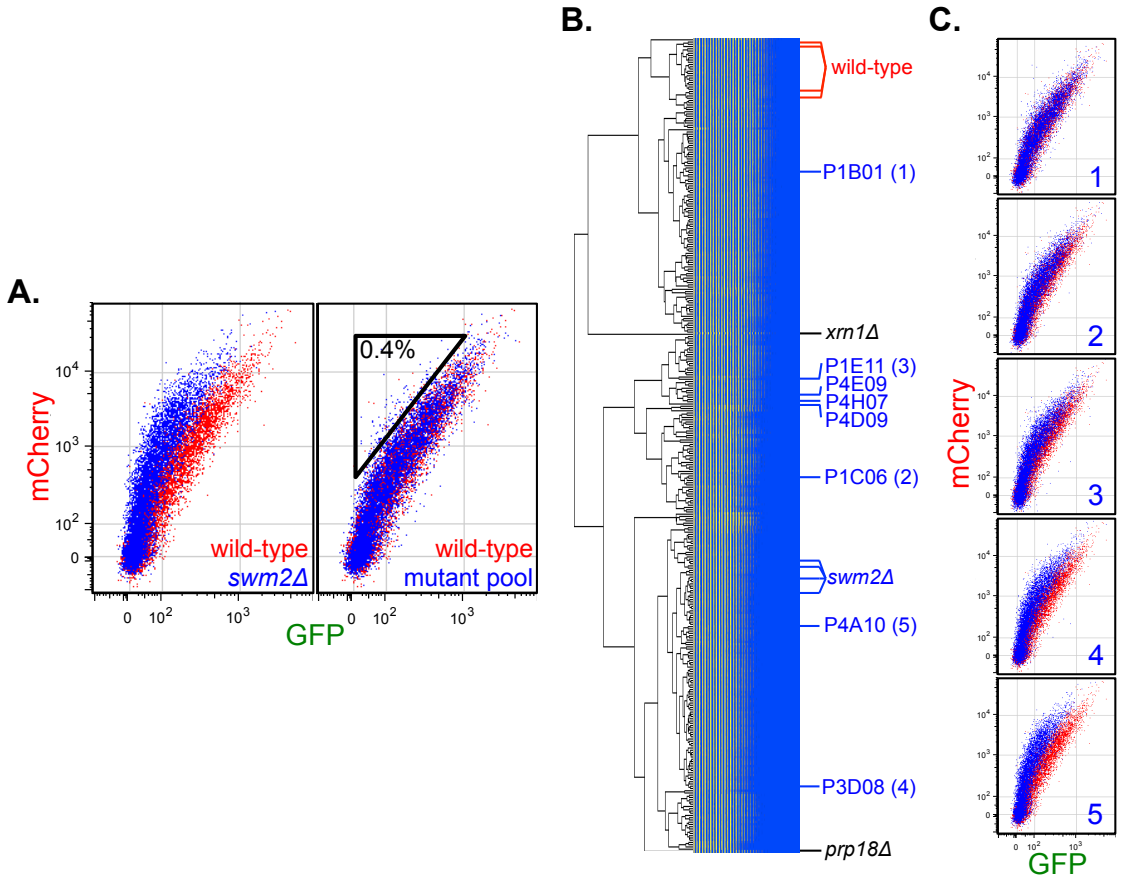


Figure 4.1: Deletion and mutation of *SWM2* results in quantifiable reporter phenotypes.

(A) Phenograph overlay is shown with wild-type (red) cells in the background and either *swm2Δ* or pooled mutant cells (blue) in the foreground. All cells are expressing the gene expression reporter. The black open triangle in the right panel indicates the gate used to sort potential *swm2* mutants of interest. 8000 cells are shown for each dataset. (B) Binning and clustering analysis of reporter flow cytometry data from several hundred *swm2* mutants and controls (wild-type, *swm2Δ*, *xrn1Δ,* and *prp18Δ*). Controls and clones of interest are labeled. Numbers in parentheses refers to the class of *swm2* mutant and these samples are visualized in panel C. (C) Representative sample phenographs from the five arbitrary classes of mutants (labeled with blue number), 1 having little to no effect and 5 being similar to *swm2Δ*.

4.3.2 Analysis, validation, and characterization of *swm2* mutants

To determine which of the sorted clones had a reproducible and significant effect on reporter expression, we isolated individual clones and stored them in specific wells within 96-well plates, alongside controls of wild-type *SWM2* and empty vector (*swm2Δ*). Several hundred clones were selected and analyzed via high-throughput flow cytometry in this secondary screen. We were able to classify the mutants into five arbitrary classes; class 1 having no effect, up to 5 having a phenotype similar to *swm2Δ* cells. Clones were evenly dispersed between these groups and we were encouraged that many had a significant effect on reporter expression. To analyze these several hundred clones in a more systematic manner, we used our binning and clustering algorithm, CHUBACA, resulting in a dendrogram that essentially grouped clones with similar phenographs (Figure 2.1B). Representative phenographs from the analysis are shown in Figure 2.1C, covering the five previously described arbitrary classes of mutants.

We observed nearly one-third of the mutant clones clustering in a large group near the top that contained the wild-type controls (Figure 2.1B). Most clones in this clade were classified as 1 and 2, reflecting a minor or insignificant effect on reporter expression. The large cluster near the bottom of the dendrogram primarily contained class 4 and 5 mutants, along with the *swm2Δ* controls. The relatively smaller clade in the middle consisted of class 2 and 3 mutants that had an intermediate effect (i.e. P1E11, P4H07, and P1C06). To

monitor clustering behavior, we also incorporated data from *xrn1Δ* and *prp18Δ* cells as green- and red-shifted bootstrap controls respectively (Larimer et al. 1992; Vijayraghavan et al. 1989). Altogether we trust that our analysis effectively grouped mutant clones into groups based on the reporter phenotypes.

It was very encouraging to have many clones display significant effects on our gene expression reporter. With any random mutagenesis approach, there is the possibility of generating mutants that result in premature stop codons or destabilization of the resulting polypeptide. To quickly identify such mutants, we performed numerous western blots, probing for the C-terminal TAP tag on Swm2p (Figure 4.2). Approximately 20% of the nearly 80 sampled clones had no detectable levels of Swm2p while 30% had reduced levels. Ten non-expressing clones were sequenced and 8 contained premature stop codons while 2 were expected to produce full-length protein. Not surprisingly, the majority of the clones without detectable levels of Swm2p were previously identified as class 4 or 5 by the reporter phenotype, mimicking a *swm2Δ* situation.

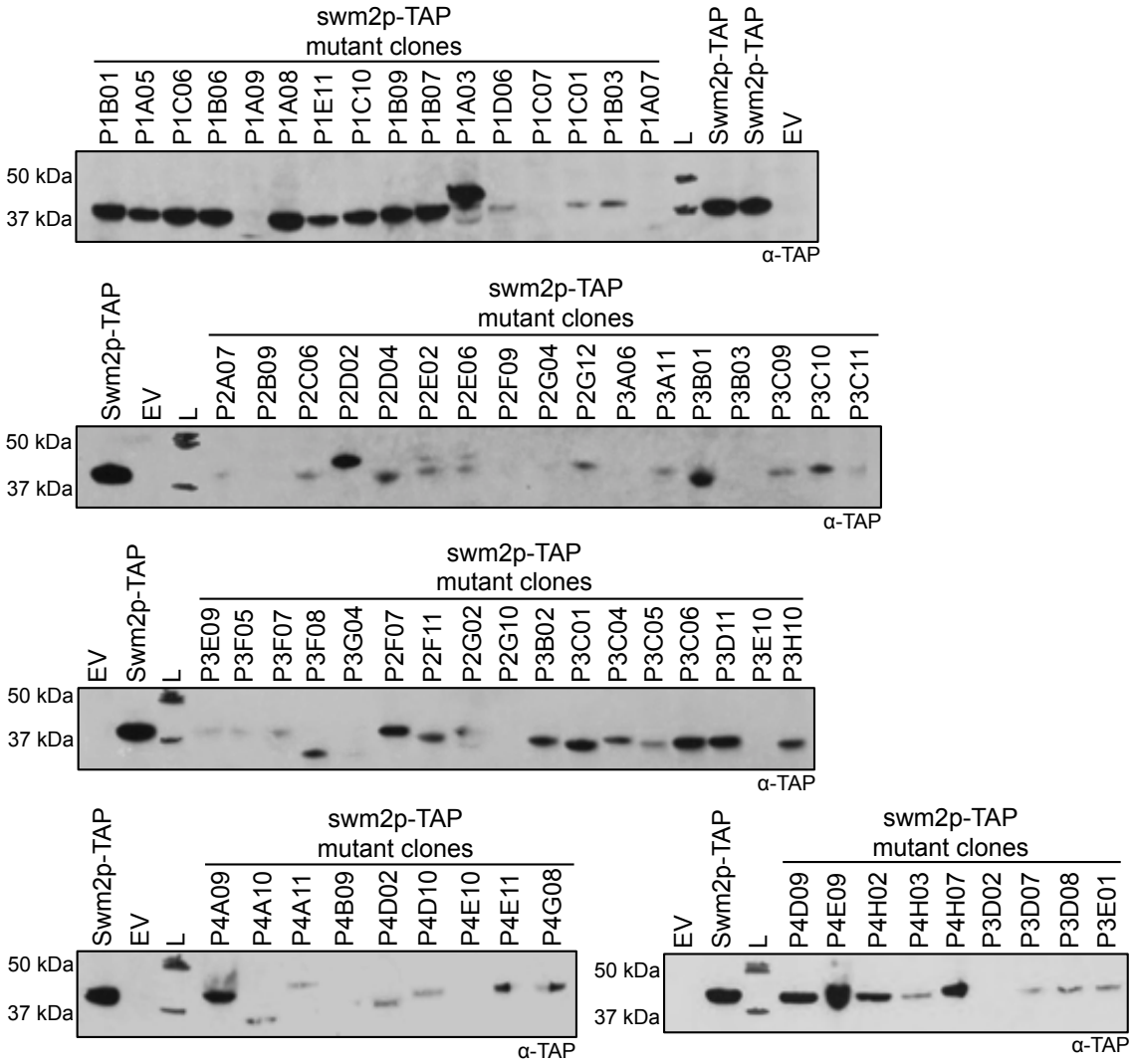


Figure 4.2: Identification of high-confidence *swm2* mutants.

(A) Western blot analysis of Swm2p-TAP expression levels. All clones, including controls, were plasmid based and in the *swm2Δ* strain. Peroxidase anti-peroxidase antibody was used to probe for the C-terminal TAP tag on Swm2p (EV, empty vector; L, Precision Plus Protein Dual Color Standard). Western blots were performed for nearly 80 clones.

For class 2 to 5 clones expressing wild-type-like levels of Swm2p, I isolated the *SWM2*-TAP plasmid and reintroduced them into the original *swm2Δ* strain with reporter and analyzed reporter expression and Swm2p levels. This served to verify that the resulting effect on reporter expression was most likely due a mutation of the plasmid borne *SWM2* and not a genomic mutation or other deleterious occurrence. Out of the 20 clones I attempted to verify, 14 retained a reporter phenotype when reintroduced into yeast (Figure 4.3A). Only 4 of the 14 expressed wild-type-like levels of Swm2p when retransformed (Figure 4.3B), which was surprising since they all expressed well in earlier rounds. To rule out the possibility that multiple *SWM2* plasmids were in these clones, additional clones from these were analyzed, but none showed increased expression (data not shown). In further analysis of the four highly expressing mutants, I found that clone P1A05 had a non-reproducible reporter phenotype and was discarded for this reason. The three stared mutants in Figure 4.3B were selected for subsequent analysis.

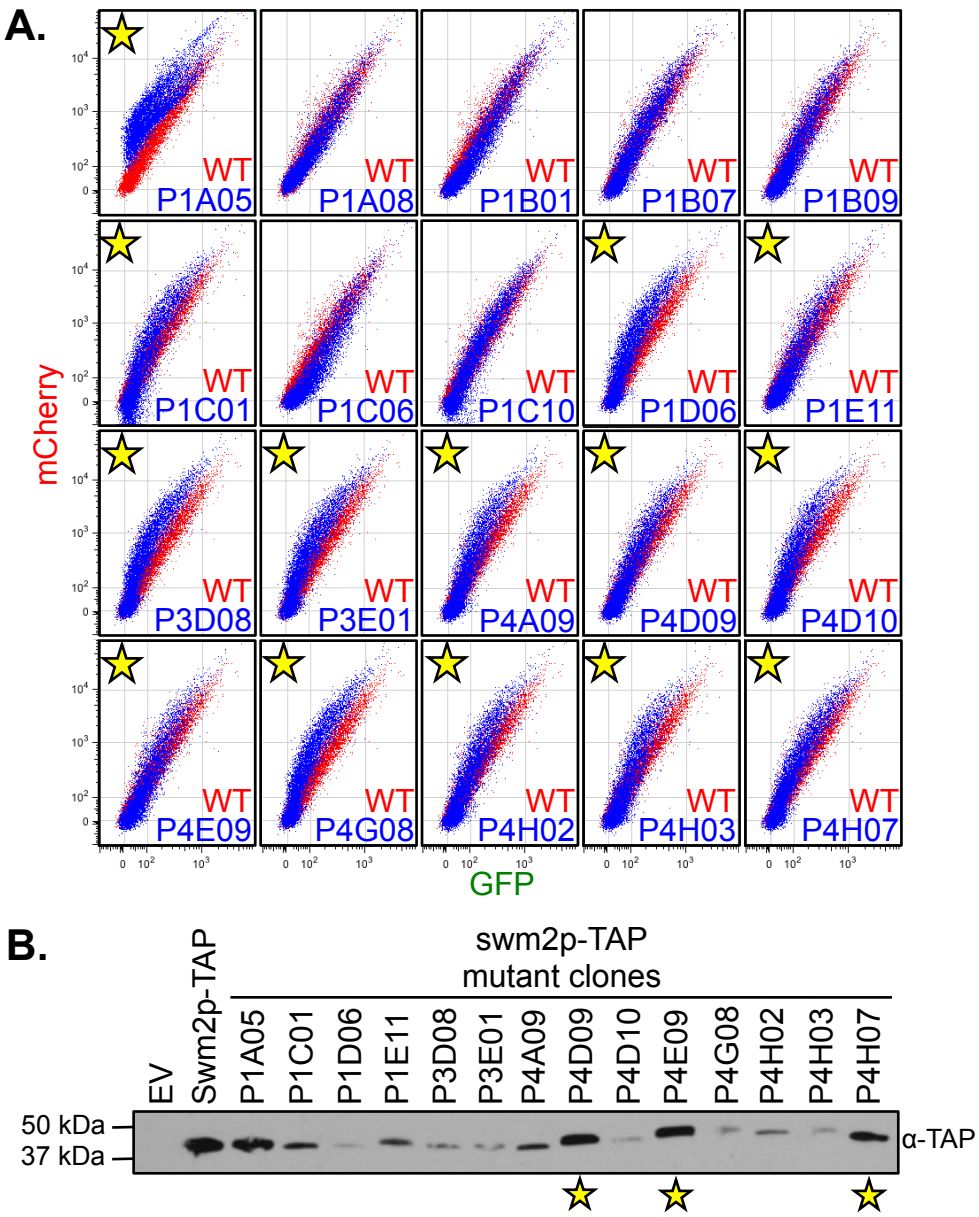


Figure 4.3: Characterization of retransformed *swm2* mutant clones.

(A) Gene expression reporter phenographs for retransformed *swm2* mutant clones (blue), overlaid onto data from wild-type cells (red). Yellow starred clones were validated in panel B. (B) Western blot analysis of Swm2p-TAP from indicated clones. Peroxidase anti-peroxidase antibody was used to probe for the TAP tag on Swm2p (EV, empty vector). Clones with yellow stars indicate clones that were highly expressed and thus validated as high confidence mutants.

In Figure 4.4A, the mutations for the three high-confidence hits are mapped to the primary structure of Swm2p for the three verified clones, P4D09, P4E09, and P4H07, which had 5, 3, and 6 amino acid substitutions respectively. Reporter phenograph overlays for these mutants are shown in Figure 4.4B and retain a class 2 or 3 like phenotype. Although there were more mutations than we had hoped for, we were encouraged that half of these mutations were in the C-terminal portion of Swm2p. This agrees with reporter expression data from C-terminal truncations of Swm2p, in which I noticed increasing mCherry/GFP ratios as subsequent 15 amino acid portions are deleted, reaching a similar ratio as *swm2Δ* between truncations of 30 and 45 amino acids (Figure 4.5). Of note, truncations of more than 45 amino acids result in drastically reduced or abolished signal when soluble proteins are analyzed by western blot (Figure 4.5C; Yaffe and Schatz 1984).

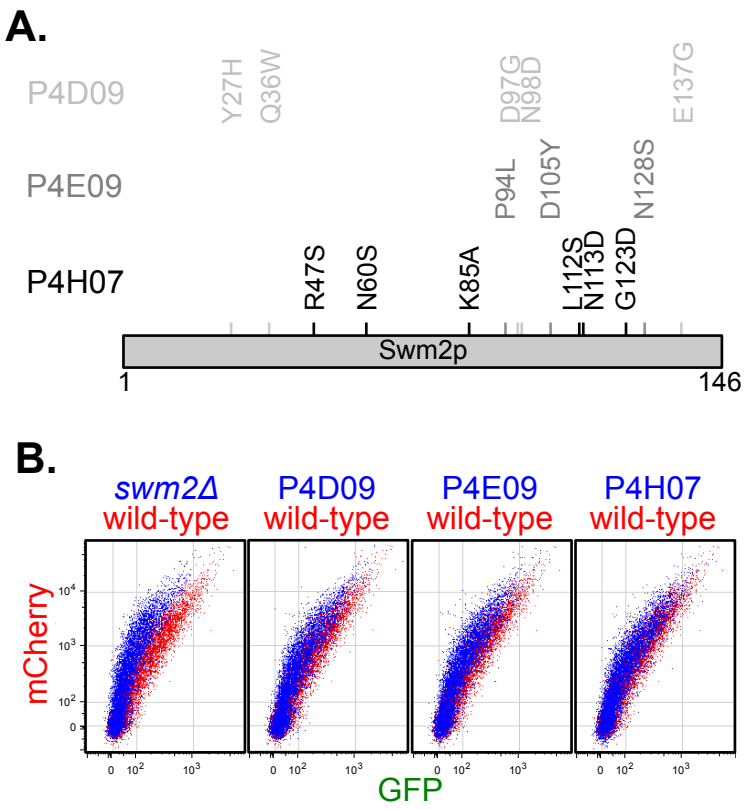


Figure 4.4: Amino acid substitutions in the three high-confidence *swm2* mutants cluster towards the C-terminus of Swm2p.

(A) Mapping point mutations in the three verified clones P4D09, P4E09, and P4H07. The three rows contain the mutations specific to the clone labeled to the left and color-coded while specific positions are depicted on the top of the cartoon and color-coded. (B) Gene expression reporter phenographs for *swm2Δ* and the three verified high-confidence mutants (blue), overlaid onto data from wild-type cells (red).

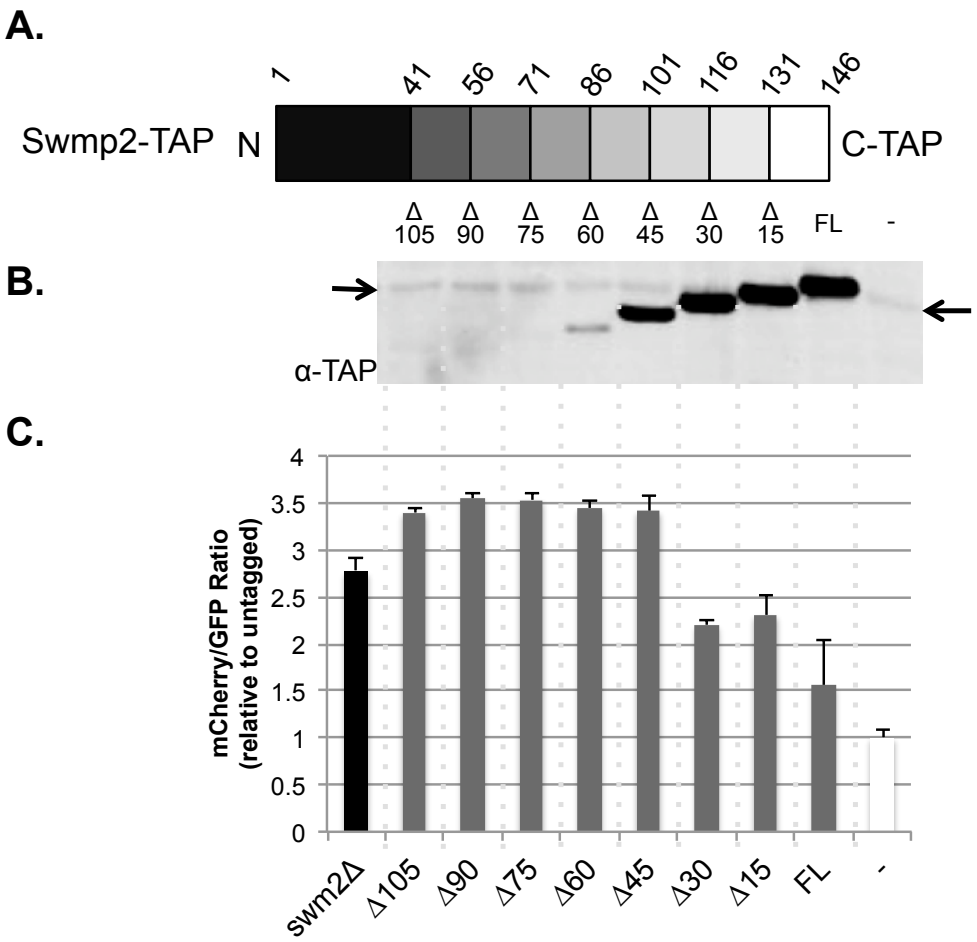


Figure 4.5: C-terminal truncations of Swm2p result in an increase of the mCherry/GFP reporter signal ratio.

(A) The cartoon is depicting the truncations made by genomic C-terminal TAP tags on Swm2p. The amino acid position/numbers are labeled on the top while the C-terminal truncation size is indicated below. (B) Western blot analysis of Swm2p C-terminal truncations. Labels continue from panel A (FL, full length Swm2p-TAP; -, untagged BY4741). Peroxidase anti-peroxidase antibody was used to probe for the TAP tag on Swm2p. Nonspecific background band is indicated with the small black arrow. (C) Reporter expression ratios of Swm2p truncations and deletions. Reporter expression values (GFP and mCherry) were determined by taking the mean values for the 21,000 cells analyzed. The mCherry/GFP ratio (pre-mRNA/mRNA) was calculated and made relative to the untagged strain. Error bars represent the standard deviation of three biological replicates. Dashed lines are used to align the samples with the western blot for interpretation.

4.3.3 Mutations of Swm2p result in gene-specific pre-mRNA splicing defects

I identified three mutants with a reproducible red-shifted reporter phenotype, but wanted to determine if there were splicing defects of endogenous introns. To reach this goal, I performed RT-qPCR analysis of the splicing efficiency of four endogenous introns using total RNA from wild type, *swm2Δ*, and *swm2* mutant strains. When analyzing the data, I observed gene-specific effects on pre-mRNA splicing (Figure 4.6). For *RPS21B,* all three *swm2* mutants have elevated pre-mRNA/total ratios that are similar to the knockout mutants (*swm2Δ*). In contrast, for *SRB2* and *TEF5* splicing, the mutants have muted yet fairly similar splicing defects to each other when compared to *swm2Δ*. Standing out is *DBP2,* in which I detect *swm2Δ*-like splicing defects for *swm2-P4D9* and *swm2-P4H7*, yet no measurable splicing defect for *swm2-P4E9*. It is very interesting that we observe such a diverse range of splicing phenotypes for the three mutants across each of these introns. I attempted to dissect the three mutants in order to find the individual point mutants or groups of mutants that contributed to the splicing defects observed, but none presented comparable effects (data not shown). We chose to continue characterizing the three parental mutants based on these findings.

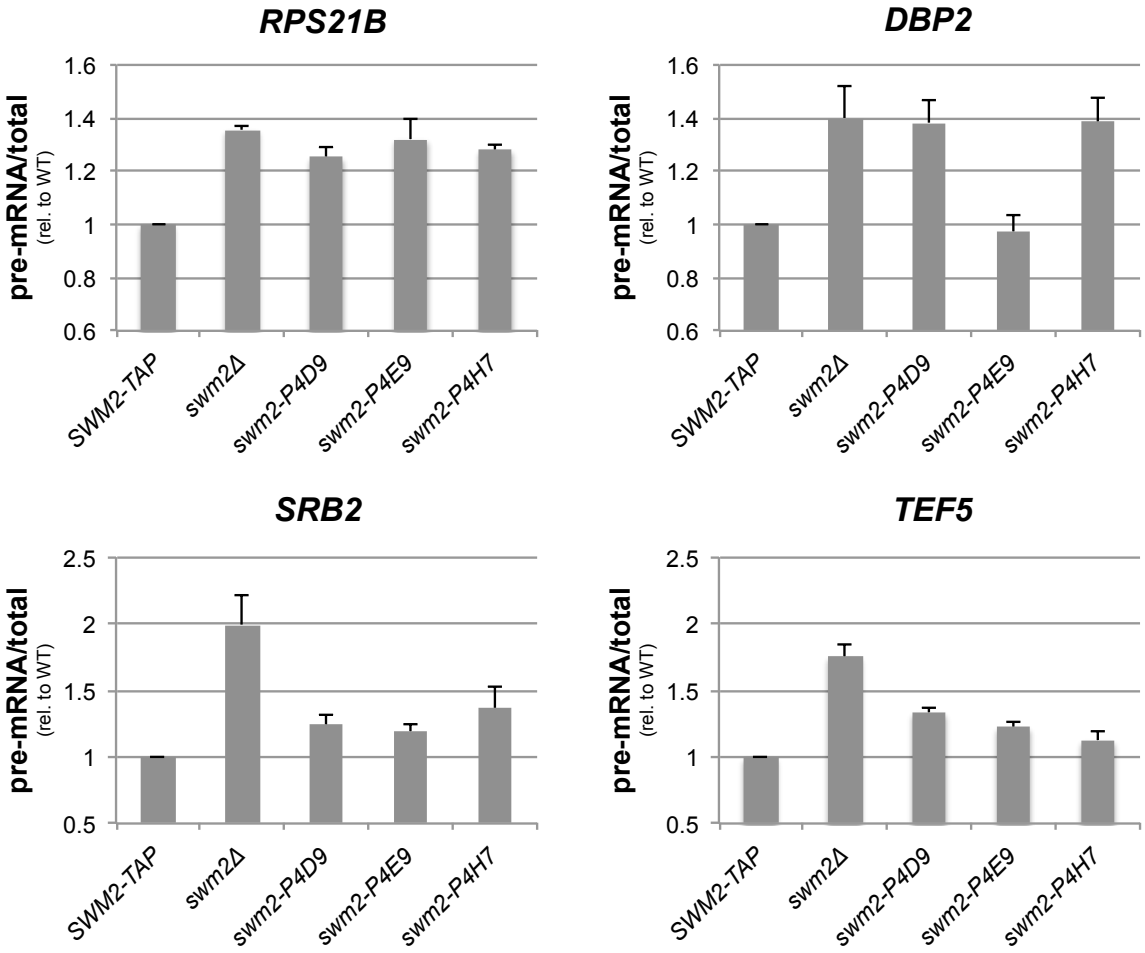


Figure 4.6: *swm2* mutants exhibit gene-specific pre-mRNA splicing defects.

RT-qPCR analysis was performed using total RNA from the indicated strains (SS4159-4163) for the specified intron/gene. Primers for qPCR are described in Table 4.3. The quantities of pre-mRNA and total transcripts were quantified using a standard curve and ratios are relative to the wild-type control (*SWM2*-TAP). All strains are based off the *swm2Δ* strain with pRS415-derived vectors and grown in synthetic medium lacking leucine. Error bars represent the standard error between biological triplicates.

In identifying gene-specific pre-mRNA splicing defects in the *swm2* mutant strains, we were interested in knowing if these mutants had any affect on growth rate. Although the *swm2Δ* strain does not have any obvious growth defects, it does have a synthetic growth defect with *mud2Δ* (Chang et al. 2010), which codes for a U1 snRNP component (Liao et al. 1993; Abovich et al. 1994; Rutz and Séraphin 1999). This synthetic growth phenotype is also seen when Swm2p is lacking from strains with mutations of the splicing factors Brr1p, Cwc21p, Dbr1p, Mud2p, Prp42p, Snu56p, and Snu66p (Wilmes et al. 2008; Chang et al. 2010). To test these three mutants for synthetic growth defects, I first verified the *mud2Δ swm2Δ* synthetic growth defect that had been previously reported. As seen in Figure 4.7A, the *mud2Δ swm2Δ* strain has a synthetic growth defect at low, permissive and high temperatures when compared to the single-mutant and wild-type cells, with the largest difference at the elevated temperature. When empty, wild-type, and the mutant *swm2* plasmids were introduced into the *mud2Δ swm2Δ* strain and tested on minimal media at the same temperatures, there were no discernible growth defects for the double deletion or any of our mutants combined with *mud2Δ* (Figure 4.7B). It does appear that the four strains with a version of *SWM2* slightly complement growth at 37°C.

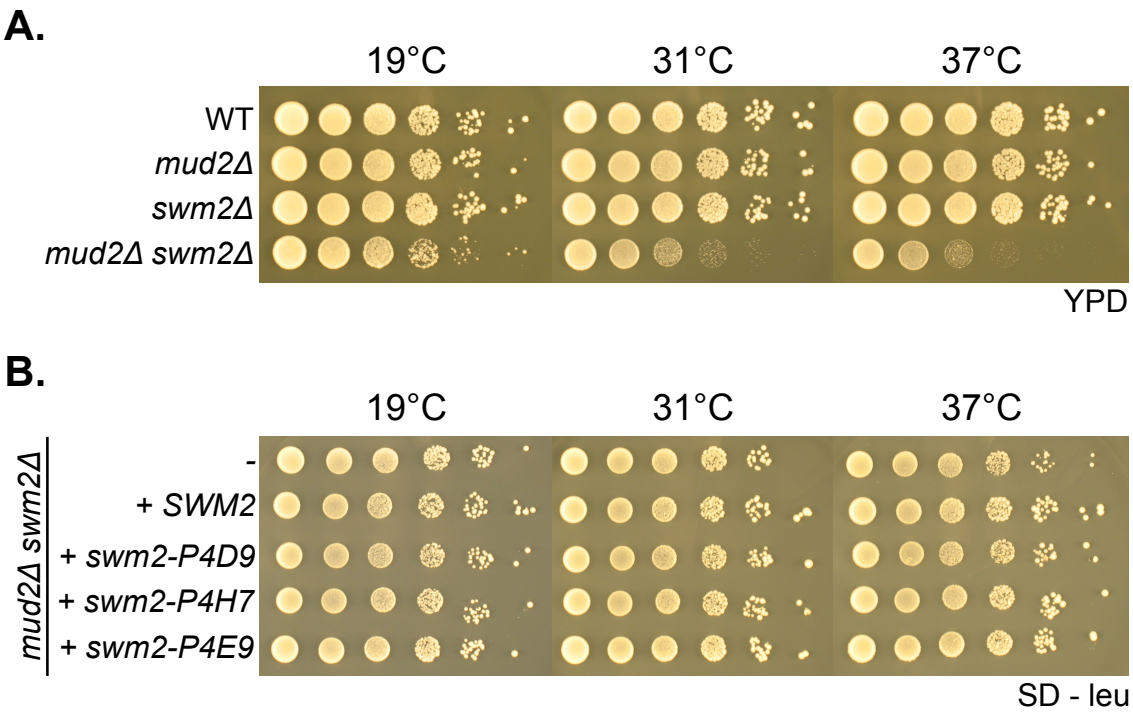


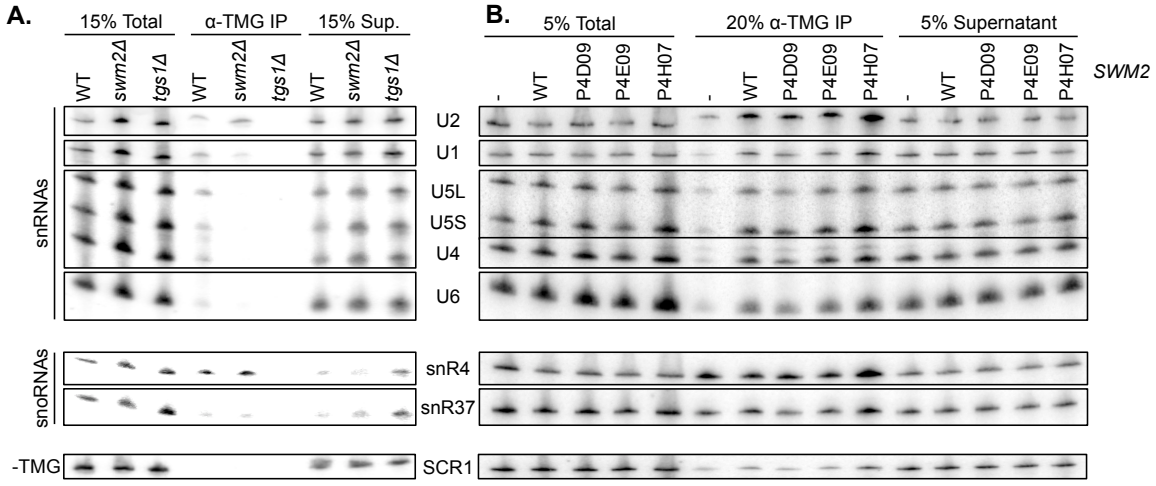
Figure 4.7: *swm2* mutants do not exhibit synthetic growth defects with *mud2Δ* on minimal media.

Serial dilutions of the indicated strains (labeled to the left) were made and spotted onto rich (YPD) or minimal (SD – leu) media and incubated at three different temperatures (31°C is permissive). Strains SS4147-4150 were used in A and strains SS4164-4168 were used in B.

4.3.4 Partially reduced trimethylation of snRNAs in *swm2* mutants

With gene specific pre-mRNA splicing defects in our *swm2* mutants and previous reports describing reduced snRNA trimethylation in *swm2Δ* cells (Boon and Kos 2010), we wanted to test whether our mutants resulted in reduced snRNA trimethylation when compared to wild-type cells. To verify the published results and to establish the assay in our laboratory, anti-trimethylguanosine antibody was used in immunoprecipitations from extracts derived from wild-type, *swm2Δ* and *tgs1Δ* cells. As we expected, there is a complete loss of trimethylation of the spliceosomal snRNAs and snoRNAs in the cells lacking the trimethylguanosine synthase Tgs1p (*tgs1Δ*; Figure 4.8A; IP lane). Agreeing with the previously published results (Boon and Kos 2010), I also observed a specific reduction of snRNA trimethylation in *swm2Δ* cells (Figure 4.8A).

I then repeated this assay with the three *swm2* mutants and observed varying degrees of snRNA trimethylation (Figure 4.8B). In calculating IP efficiency, there are clear differential effects on snRNA trimethylation. Across the board, U5L snRNA trimethylation is reduced by about 20%. Although the mutants do not have consistent loss of specific snRNA trimethylation, they each have 2 or 3 that are significantly reduced. These results describe the steady state of snRNA trimethylation, which might hide any slight defects in snRNA maturation since spliceosomal snRNAs have a half-life of many hours (Stutz et al. 1993; Noble and Guthrie 1996).



C.

TMG IP Efficiency

	swm2Δ	*SWM2*	*swm2-P4D9*	*swm2-P4E9*	*swm2 P4H7*
U2	28.2	100	97.7	149.6	47.4
U1	16.2	100	107.3	78.8	119.8
U5L	27.6	100	81.6	79.5	73.7
U5S	27.3	100	90.6	81.9	109.9
U4	18.5	100	100.3	119.8	105.6
U6	20.8	100	89.8	101.9	108.4
SCR1	102.2	100	99.2	101.6	111.3
snR37	118.0	100	99.5	103.8	154.0
snR4	118.3	100	137.9	81.0	163.7

Figure 4.8: *swm2* mutants exhibit intermediate trimethylated-snRNA levels.

Immunoprecipitations using anti-trimethylguanosine antibody (K121) were performed with the strains labeled at the top of the lanes. The resulting RNA samples were precipitated and ran on a 7% acrylamide gel. After transfer the blots were probed for the indicated RNAs using end-labeled DNA probes (A and B; see Table 4.3 for sequences). For panel B, all strains used the *swm2Δ* background with empty vector (-), wild-type (WT) *SWM2* or mutant *swm2* (SS4159-4163). (C) IP efficiency was calculated by quantifying the respective IP bands, normalizing to the total signal, and made relative to wild-type *SWM2*.

4.3.4 Swm2p mutants have reduced interaction with Tgs1p

It was previously mentioned that Swm2p interacts with the trimethylguanosine synthase Tgs1p in yeast cells (Boon and Kos 2010). If Swm2p is acting as a specificity factor for Tgs1p by directing them to snRNAs, it is possible that the three mutants we have identified have reduced interaction with Tgs1p. To test this possibility I genomically 13myc-tagged Tgs1p in a *swm2Δ* strain. I then introduced empty vector, wild-type and mutant Swm2p-TAP vectors and prepared extract. A TAP coimmunoprecipitation was performed and western blot analysis was used to identify both the bait (Swm2p-TAP) and Tgs1p-13myc (Figure 4.9). It was obvious that the levels of bait differed between the samples so I used the bait signal to normalize the bound myc signal. When looking at the IP efficiencies, we observed that Tgs1p-myc immunoprecipitated with efficiencies of 57-65% with the mutants compared to wild-type Swm2p. These results support a model in which Swm2p interaction with Tgs1p could be important for efficient pre-mRNA splicing, although further experimentation is needed to elucidate the exact mechanism by which this might occur.

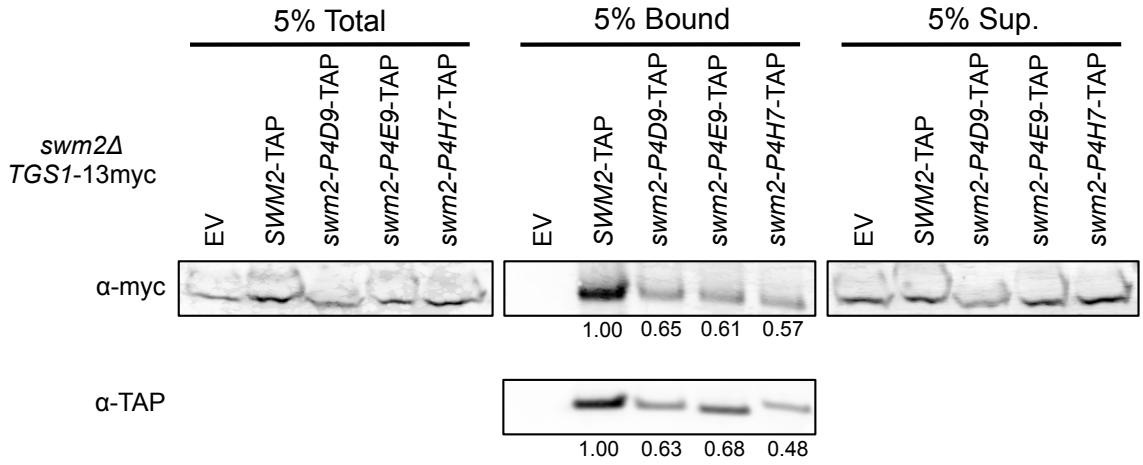


Figure 4.9: Coimmunoprecipitations reveal decreased affinity of Swm2p mutants for Tgs1p.

TAP coimmunoprecipitations were performed using IgG sepharose and extract from a *swm2Δ* Tgs1p-13myc strain with the indicated vectors (EV, empty vector; strains SS4154-4158). Protein samples were separated in a SDS-PAGE gel and transferred to nitrocellulose. Western blot analysis was performed with anti-myc and anti-TAP antibodies and chemifluorescence reagents. All blots were performed on the same membrane, separated to remove control lanes and adjust exposure. The levels of bait (Swm2p-TAP) were quantified and normalized (to wild-type) levels are reported below the anti-TAP blot. The myc signals for Tgs1p-myc were then quantified. Immunoprecipitation efficiency was calculated by dividing the myc signal by the bait levels, making the values relative to wild-type. The normalized IP efficiencies are reported below the anti-myc blot for the bound fraction.

4.4 DISCUSSION

When we began our inquiry into *YNR004W/SWM2,* little was known about its function, except that it had a similar genetic interaction profile to *TGS1* and that its deletion resulted in a similar splicing phenotype to *tgs1Δ* at decreased temperature when using splicing-specific microarrays (Wilmes et al. 2008). Additionally, high-throughput protein interactions studies identified Swm2p and Tgs1p as interacting partners (Uetz et al. 2000; Yu et al. 2008). With no recognizable motifs or domains, we decided to perform a random mutagenesis screen to identify mutants that affected its function. By combining the standard gap-repair method with our reporter and fluorescence-activated cell sorting, we quickly identified three mutants of Swm2p with multiple point mutations that elicited gene-specific splicing defects of endogenous introns (Figure 4.6). Additionally, these three mutants appear to have significantly reduced interaction with Tgs1p (Figure 4.9), and slightly reduced levels of trimethylated snRNAs (Figure 4.8).

As discussed earlier, the relatively small effect on snRNA trimethylation might not be an accurate representation of Swm2p's importance in this role since snRNAs are believed to be very stable (Stutz et al. 1993; Noble and Guthrie 1996). A significant effect on snRNA trimethylation could be masked by increased stability of the trimethylated species compared to the unmodified species. One way to look at this more closely would be to combine a pulse-chase type of approach with the TMG immunoprecipitation experiment to query the

mutant's effect on *de novo* snRNA trimethylation using labeled S-adenosyl methionine. Additionally, the TMG immunopreciipations and Swm2p/Tgs1p coimmunoprecipitations are in the process of being repeated multiple times to determine the reproduciblilty of these results.

In the course of characterizing the *swm2* mutants from the screen, one problem I encountered was a large number of clones that were not expressing wild-type levels of Swm2p when using western blot techniques to screen them. Although destabilizing and truncation mutants are a common occurrence in mutagenesis screens such as ours, it would be advantageous to quickly eliminate or flag these mutants. One possible solution would be to replace the TAP tagged I used with a fluorescence tag that is distinguishable from GFP and mCherry by flow cytometry. The researcher could then gate the cells based on this initial fluorescence parameter to identify clones that are adequately expressing the protein of interest. This method could radically decrease the number of false positive hits produced in this type of screen. In screening over 10 million cells in the FACS sort, we were not surprised that a handful of clones did not result in a reporter phenotype when tested further. With this number of clones, we could have inadvertently isolated genomic mutations that resulted in a red-shift.

When dissecting the individual point mutants and groups of point mutants from the three isolated mutants, I did not observe phenotypes similar to the parental/original mutants. Since I observed both the amino acid substitutions

clustering towards the C-terminus of the protein (Figure 4.4A), and the detrimental effect of truncating the C-terminus of Swm2p (Figure 4.5C), it would be potentially advantageous to start at the C-terminus and combine the individual point mutants. I would predict that this combined mutant, which would include multiple charge changes, would interfere with protein-protein interactions.

One unexpected result was the lack of synthetic growth defects on minimal media with the *mud2Δ swm2Δ* (Figure 4.7B). This was surprising because we verified the strong synthetic growth defect on rich media (Figure 4.7A). Even though the *SWM2* and *swm2* mutant plasmid appear to slightly complement the double deletion at elevated temperatures, I expected to see a much larger difference. The strong negative genetic interaction could also be a great tool to screen future mutants, if we are able to develop a working situation to observe growth phenotypes. Additionally, it would be advantageous to use such a situation to also test the *swm2* mutants for genetic interactions with splicing factor deletions or conditional splicing mutations to further discern the function of Swm2p.

Since it appears that Swm2p is instrumental in snRNA trimethylation, I tested whether Swm2p-TAP can isolate snRNAs. Compared to a positive control of the snRNP component Smb1p, Swm2p and mutants purified near background levels of snRNAs (data not shown). This would be worth testing again since there was excessive background. Moreover, it would be worthwhile to see what additional proteins, besides Tgs1p, that Swm2p interacts with using mass

spectrometry or a candidate approach with snRNA biogenesis or splicing factors. With Swm2p's function in hypermethylation specificity, it is likely that its interaction with Tgs1p is transient and not stable. Additionally, Tgs1p has been shown to isolate U4 and U5 snRNAs (Mouaikel et al. 2002, 2003), and it would be worthwhile to test this interaction in the context of our mutants to determine if Swm2p mediates this interaction. Since the nucleolar localization of Swm2p is dependent on Tgs1p, and not vice versa (Boon and Kos 2010), it is expected that our mutants might be mislocalized since they have reduced Tgs1p.

Altogether our data support a model in which Swm2p interacts with Tgs1p and directs hypermethylation of snRNA caps. One outstanding question is how Swm2p specifically targets Tgs1p to the snRNAs. Since I did observe above background levels of snRNAs immunoprecipitating with Swm2p, it is possible that Swm2p directly or indirectly binds to snRNAs or snRNPs. As there are no reported physical interaction candidates, we could directly test for physical interaction with snRNP components, potentially focusing on splicing factor genes with genetic interactions with *SWM2* (many negative interactions). It is interesting that the limited number of positive genetic interactions are mostly genes coding for nucleolar localized components, or genes involved in rRNA biogenesis, mRNA export or mRNA degradation. The most striking negative genetic interaction that has been reported for *SWM2* is *MUD2*, which codes for a protein that directly binds the intron branchpoint and interacts with U1 snRNP early on in spliceosome formation (Abovich et al. 1994; Rutz and Séraphin 1999). It is

possible that Swm2p has a role with Mud2p or trimethylation of snRNAs is particularly important when Mud2p is absent, as a *mud2Δ tgs1Δ* mutant is inviable (Hausmann et al. 2008; Benarroch et al. 2009; Chang et al. 2010). Although I did not notice any particular features of the branchpoints that correlate with the gene specific splicing defects we observed (Figure 4.6). I am also in the process of testing a *tgs1Δ* mutant with our RT-qPCR assay looking at the splicing efficiency of the same introns.

I am confident that our reporter will be able to serve yeast molecular biologists and geneticists interested in applicable gene expression events. I have demonstrated that our fluorescent reporter can be used with FACS to perform rapid screens of millions of mutant clones, as compared to conventional screening tools such as differential growth on plates. Additionally, our reporter can be used as an orthogonal assay to screen previously identified mutants and conditions for gene expression effects. As I have stated earlier, it is important to test the effects on endogenous substrates and targets, as our reporter is a synthetic and unnatural gene. Although I have not described other screening efforts I have undertook, it is important to note that our reporter is tunable, as I have made it specifically sensitive to the Rnt1p endonuclease by introducing a cleavage site within the intron. It will be exciting to see our reporter technology aid other researchers in their efforts to more fully understand a diverse range of gene expression processes.

Chapter 5: Significance and Future Directions

5.1 Significance

When I began my studies in the laboratory of Scott Stevens, I was interested in improving upon the pre-mRNA splicing reporter technology that was available for the yeast *Saccharomyces cerevisiae* (Legrain and Rosbash 1989; Lesser and Guthrie 1993). Eliminating the constrains of these powerful yet aging reporter systems while taking advantages of a sensitive and high throughput technique, we successfully developed and validated a flexible single-cell two-color reporter system (Sorenson and Stevens 2014). Serendipitously our carefully designed reporter not only accurately identified *cis*- and *trans*-acting defects in pre-mRNA splicing, but it also was sensitive and specific to defects in other gene expression processes including transcription, chromatin dynamics, mRNA export and mRNA degradation/surveillance.

The high-throughput nature of our reporter system was exemplified in our successful screening of the haploid deletion collection, consisting of nearly 5000 unique deletion mutants. With this screen we were able to identify an uncharacterized gene, *SWM2/YNR004W,* that was partially characterized while our efforts were underway (Boon and Kos 2010; Chang et al. 2010). This rich dataset, along with our binning and clustering analysis pipeline CHUBACA, provide a valuable reference and tool for comparing novel genetic or chemical perturbations. Standard genetic screens using budding yeast commonly utilize tedious and resource consuming colony-based assays to identify mutants of

interest. Utilizing our reporter system a researcher can explore a number of different gene expression processes rapidly utilizing fluorescence-activated cell sorting to accelerate the identification of interesting mutants from complex populations. Furthermore the modification of our assay for small molecule and cyclic peptide screening identified two molecules of high-interest.

5.2 FUTURE DIRECTIONS

This dissertation is not by any means an exhaustive search of genes and compounds, but serves to highlight the development, validation and utilization of our reporter system. One significant project that resulted from screening a large collection of synthetic histone mutants resulted in the identification a specific histone modification that impacts pre-mRNA splicing efficiency of our reporter and endogenous introns. In collaborating with two different laboratories, we were unable to identify a specific mechanism by which this connection occurs. Excitingly, we discovered that deleting the enzyme responsible for this histone modification fully suppresses the splicing defect caused by the deletion of a specific splicing factor. This specific phenomenon and the general mechanisms behind the connections between pre-mRNA splicing and histone modification are worth exploring further.

5.2.1 Small molecule screening

I believe that the most fruitful avenue to take with the reporter is to continue and expand the small molecule screening efforts. The few thousand compounds we tested in our pilot screen are a mere sliver of the number of compounds available in large-scale screens. I am confident that we were able to successfully adapt our reporter for 96-well plate reader analysis or even 384-well plates in a drug-sensitive strain (Suzuki et al. 2011). Additionally, with the newly installed high-throughput flow cytometry instrument, screens by flow cytometry can be performed an order of magnitude faster that the previous setup (Edwards et al. 2001, 2007; Black et al. 2011). If one wanted to perform multiple orthogonal assays or screen with our reporter in multiple yeast strains simultaneously, a multiplexing approach could be taken (Chen et al. 2012; Kulak and Lum 2013). Last, it would be advisable to further study the pre-mRNA splicing defect caused by TI3D_0104_B19.

5.2.2 Swm2p function

In regards to the mutations we identified of Swm2p that results in gene-specific pre-mRNA splicing defects and decreased interaction with the trimethylguanosine synthase Tgs1p, there are several experiments that should be completed. First, repeating the trimethylguanosine immunoprecipations and Swm2p coimmunoprecipitations will determine snRNA trimethylation levels and Tgs1p interaction with more certainty. Since Tgs1p is required for Swm2p's

nucleolar localization (Boon and Kos 2010), it would be interesting to determine if the three mutants described in Chapter 4 have reduced nucleolar residency. Last, it would be interesting to combine individual point mutations contained in the three isolated mutants, focusing on the C-terminal substitutions that involve a change in charge. Given the apparent importance of the C-terminus, this could lead to a more complete loss-of-function mutant as compared to the ones already isolated.

5.2.3 Adjusting reporter sensitivity

Another aspect of our reporter that can be taken advantage of is that one can modify the sensitivity of the reporter by addition of *cis*-acting elements. In collaboration with Guillaume Chanfreau at UCLA, I added a Rnt1p cleavage site into the intron of our reporter, making the resulting fluorescent signals sensitive to this endonuclease. In screening an overexpression library by FACS, we were able to identify genes that resulted in altered stability of the reporter pre-mRNA when overexpressed, although they did not significantly affect the stability of endogenous targets. An alternative Rnt1p cleavage site could be tested in the future, as well as testing other RNA elements to affect the stability of the reporter pre-mRNA and/or mRNA.

5.2.4 Genes and pathways of interest

One set of genes that would be interesting to explore utilizing our reporter to study in more detail are those coding for TREX-2 complex components; *SAC3, THP1, SEM1, SUS1* and *CDC31*; functioning to link mRNA transcription and export (Fischer et al. 2002; Rodriguez-Navarro et al. 2004). One member of the human TREX-2 complex, GANP, is involved in cancer and is an appreciated drug target (Fujimura et al. 2005; Kageshita et al. 2006; Wickramasinghe et al. 2010; Phimsen et al. 2012). Recently TREX-2 components have been implicated in the DNA damage response, genome instability and R-loop formation via the tumor suppressor BRCA2 (Bhatia et al. 2014). The unique signal in TREX-2/mRNA export mutants, as well as the medical relevance of the human TREX-2 complex, make this is an ideal area of future study with our reporter.

5.3 FINAL CONCLUSION

Altogether, this work described my efforts in developing a versatile reporter system for budding yeast. Not only is our reporter a significant improvement to existing pre-mRNA splicing reporters, it enables researchers to study many different gene expression processes simultaneously, as well as the interconnectedness between these processes. The cyclic peptide, small molecule, and *SWM2* high-throughput screens are prime examples of the utility of our reporter and analysis pipeline. Moreover, the massive dataset that resulted from the yeast deletion collection provides our lab and others a prime reference

tool for the comparative analysis of additional genetic, chemical, or environmental perturbations.

APPENDIX I: ANNOTATED SEQUENCE OF GENE EXPRESSION REPORTER

GCGGCCGCGAATTCACTAGTGATTCTCGAGTAACTTTATTTAGTCAAA
AAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCG
GGTTACACAGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCT
GGCATCCACTAAATATAATGGAGCCCGCTTTTTAAGCTGGCATCCAGAAAAA
AAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAGTTCATAG
GTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGCAAAAA
ACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGAC
ACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTATT
ACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAG
TTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTA
TTGATTGTAATTCCGTAAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATA
GTTAGTCTTTTTTTTAGTTTTAAAACACCAAGAACTTAGTTTCGAATAAACACA
CATAAACAAACAAAGGATCCATGTCTTCCAGATTCACTAAGACTAGAAAGCA
CAGAGGTCACGTCTCAGGgtATGTAGTTCCATTTGGAAGAGGGAATGAAAGA
ACCAAATATGTCGTGTGTATATCATGGTACAGCATGCTTTCTAAGGGTGAAG
AAGACAACATGGCTATCATCAAGGAATTCATGAGATTCAAGGTTCACATGGA
AGGTTCTGTTAACGGTCACGAATTCGAAATCGAAGGTGAAGGTGAAGGTAG
ACCATACGAAGGTACCCAAACCGCTAAGTTGAAGGTTACCAAGGGTGGTCC
ATTGCCATTCGCTTGGGACATCTTGTCTCCACAATTCATGTACGGTTCTAAG
GCTTACGTTAAGCACCCAGCTGACATCCCAGACTACTTGAAGTTGTCTTTCC
CAGAAGGTTTCAAGTGGGAAAGAGTTATGAACTTCGAAGACGGTGGTGTTG
TTACCGTTACCCAAGACTCTTCTTTGCAAGACGGTGAATTCATCTACAAGGT
TAAGTTGAGAGGTACCAACTTCCCATCTGACGGTCCAGTTATGCAAAAGAAG

ACTATGGGTTGGGAAGCTTCTTCTGAAAGAATGTACCCAGAAGACGGTGCT
TTGAAGGGTGAAATCAAGCAAAGATTGAAGTTGAAGGACGGTGGTCACTAC
GACGCTGAAGTTAAGACCACCTACAAGGCTAAGAAGCCAGTTCAATTGCCA
GGTGCTTACAACGTTAACATCAAGTTGGACATCACCTCTCACAACGAAGACT
ACACCATCGTTGAACAATACGAAAGAGCTGAAGGTAGACACTCTACCGGTG
GTATGGACGAATTGTACACTTAAGAGTTACCACGTTTCTTTTGTTTCGATAAA
ATGTCCAGTTGAAAACCTGTTTTACTAACGATTTAAAAATTGTATTTCATTACA
ATATTTTTTTTGTACagCCGGTTAAGGTCGTATCGGAGATCCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC
TGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCC
GCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA
CGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATG
GCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAAC
ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCC
CATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCC
AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
CAAGAGATCCTAAGCTGGTTCTAACTGGAAATAATTTCCATTAGATTCCTCTT
TTTCTCGTCCATTAACCAAAATATATTATTGAATTCAGCGGTTCCTTTTTTCTC
ATTTTCGCATATAGCTGCACTATTAGAATCAGCCCACTCTAGGTAAACACAG

TTCCTCGATATACCTCTGTCTTACTATCAGTGGTTAAACCTTATGCAAATATA
ATATATATATATATATATATATATATATATCTCATACTTTTGTTGATTCTTGTGTAA
TTATTGGAAAAGACAAAACAAAGCAAGCGTTTCTATTCATATTTACAAGTATT
TTTTATGACAAACTATTTCTTAATTTTCCCACCGGCGGCTTTGAATAAGGCAA
TGTGTCGAC

APPENDIX II: TABLE OF RED- AND GREEN-SHIFTED DELETION MUTANTS

Green-shifted	Red-shifted	Red-shifted	Red-shifted
ACE2 / YLR131C	AAT2 / YLR027C	MAK32 / YCR019W	STE5 / YDR103W
ARC18 / YLR370C	AIM43 / YPL099C	MCK1 / YNL307C	SWC3 / YAL011W
ARG82 / YDR173C	APM4 / YOL062C	MET6 / YER091C	SWT21 / YNL187W
CCR4 / YAL021C	ARP6 / YLR085C	MGR2 / YPL098C	TAF14 / YPL129W
DEF1 / YKL054C	ATG14 / YBR128C	MRP51 / YPL118W	TAL1 / YLR354C
DHH1 / YDL160C	ATG21 / YPL100W	MRPL7 / YDR237W	TIS11 / YLR136C
EFG1 / YGR271C-A	BCK1 / YJL095W	MTC7 / YEL033W	TPO5 / YKL174C
GBP2 / YCL011C	BDF1 / YLR399C	MUD2 / YKL074C	UBA4 / YHR111W
GSH1 / YJL101C	BEM4 / YPL161C	NAM7 / YMR080C	UME6 / YDR207C
LTV1 / YKL143W	BUB3 / YOR026W	NMD2 / YHR077C	UPF3 / YGR072W
MFT1 / YML062C	BUD13 / YGL174W	NQM1 / YGR043C	VBA3 / YCL069W
MLP1 / YKR095W	BUD16 / YEL029C	PAC10 / YGR078C	VPS71 / YML041C
MMP1 / YLL061W	BUD32 / YGR262C	PCL8 / YPL219W	VPS72 / YDR485C
NEW1 / YPL226W	CAF40 / YNL288W	POC4 / YPL144W	VPS72 / YDR485C
NUP120 / YKL057C	CKI1 / YLR133W	POL4 / YCR014C	YBR174C / YBR174C
NUP60 / YAR002W	CNB1 / YKL190W	PRD1 / YCL057W	YCR001W / YCR001W
PLC1 / YPL268W	COX20 / YDR231C	QCR8 / YJL166W	YDJ1 / YNL064C
POP2 / YNR052C	CRG1 / YHR209W	RAV2 / YDR202C	YDR133C / YDR133C
RPS22A / YJL190C	CSE2 / YNR010W	RER1 / YCL001W	YDR239C / YDR239C
RRP6 / YOR001W	DGR2 / YKL121W	RIM1 / YCR028C-A	YEA4 / YEL004W
SAC3 / YDR159W	DPB4 / YDR121W	RPD3 / YNL330C	YGR051C / YGR051C
SEM1 / YDR363W-A	DST1 / YGL043W	RPL34A / YER056C-A	YGR071C / YGR071C
SUS1 / YBR111W-A	EAF1 / YDR359C	RPL42B / YHR141C	YHI9 / YHR029C
THP1 / YOL072W	EBS1 / YDR206W	RPN10 / YHR200W	YJL169W / YJL169W
THP2 / YHR167W	EEB1 / YPL095C	RPN4 / YDL020C	YKL102C / YKL102C
UBR2 / YLR024C	ELP4 / YPL101W	RRT2 / YBR246W	YKL171W / YKL171W
XRN1 / YGL173C	FAB1 / YFR019W	RSN1 / YMR266W	YLR358C / YLR358C
YGR031W / YGR031W	FIR1 / YER032W	RTF1 / YGL244W	YML081W / YML081W
YNL140C / YNL140C	FMP25 / YLR077W	RTN1 / YDR233C	YMR027W / YMR027W
	HOS3 / YPL116W	RTR1 / YER139C	YNL170W / YNL170W
	HTZ1 / YOL012C	SCM4 / YGR049W	YNR004W / YNR004W
	ILV6 / YCL009C	SEF1 / YBL066C	YOR305W / YOR305W
	IRC15 / YPL017C	SEO1 / YAL067C	YOR309C / YOR309C
	IRC19 / YLL033W	SET2 / YJL168C	YOR352W / YOR352W
	IRC25 / YLR021W	SGO1 / YOR073W	YPL107W / YPL107W
	IST3 / YIR005W	SNU66 / YOR308C	YPL150W / YPL150W
	ISW2 / YOR304W	SOH1 / YGL127C	YPT7 / YML001W
	IWR1 / YDL115C	SPT8 / YLR055C	ZRT2 / YLR130C
	LEA1 / YPL213W	SRB2 / YHR041C	
	LYS4 / YDR234W	SRB5 / YGR104C	

References

Abovich N, Liao XLC, Rosbash M. 1994. The yeast Mud2 protein - an interaction with Prp11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev* **8**: 843–854.

Aboye TL, Camarero JA. 2012. Biological Synthesis of Circular Polypeptides. *J Biol Chem* **287**: 27026–27032.

Albertazzi L, Arosio D, Marchetti L, Ricci F, Beltram F. 2009. Quantitative FRET analysis with the EGFP-mCherry fluorescent protein pair. *Photochem Photobiol* **85**: 287–97.

Ares M, Grate L, Pauling MH. 1999. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **6**: 1138–1139.

Arita Y, Nishimura S, Matsuyama A, Yashiroda Y, Usui T, Boone C, Yoshida M. 2011. Microarray-based target identification using drug hypersensitive fission yeast expressing ORFeome. *Mol Biosyst* **7**: 1463–1472.

Astashkina A, Mann B, Grainger DW. 2012. A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacol Ther* **134**: 82–106.

Auerbach D, Arnoldo A, Bogdan B, Fetchko M, Stagljar I. 2005. Drug Discovery Using Yeast as a Model System: A Functional Genomic and Proteomic View. *Curr Proteomics* **2**: 1–13.

Austin J, Kimura RH, Woo Y-H, Camarero JA. 2010. In vivo biosynthesis of an Ala-scan library based on the cyclic peptide SFTI-1. *Amino Acids* **38**: 1313–1322.

Austin J, Wang W, Puttamadappa S, Shekhtman A, Camarero JA. 2009. Biosynthesis and biological screening of a genetically-encoded library based on the cyclotide MCoTI-I. *Chembiochem Eur J Chem Biol* **10**: 2663–2670.

Bancos I, Bida JP, Tian D, Bundrick M, John K, Holte MN, Her YF, Evans D, Saenz DT, Poeschla EM, et al. 2013. High-Throughput Screening for Growth Inhibitors Using a Yeast Model of Familial Paraganglioma. *PLoS ONE* **8**: e56827.

Barberis A, Gunde T, Berset C, Audetat S, Lüthi U. 2005. Yeast as a screening tool. *Drug Discov Today Technol* **2**: 187–192.

Barnes WM. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci USA* **91**: 2216–2220.

Beckman RA, Mildvan AS, Loeb LA. 1985. On the fidelity of DNA replication: manganese mutagenesis in vitro. *Biochemistry (Mosc)* **24**: 5810–5817.

Beelman CA, Stevens A, Caponigro G, LaGrandeur TE, Hatfield L, Fortner DM, Parker R. 1996. An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature* **382**: 642–646.

Benarroch D, Jankowska-Anyszka M, Stepinski J, Darzynkiewicz E, Shuman S. 2010. Cap analog substrates reveal three clades of cap guanine-N2 methyltransferases with distinct methyl acceptor specificities. *RNA* **16**: 211–220.

Benarroch D, Qiu ZR, Schwer B, Shuman S. 2009. Characterization of a mimivirus RNA cap guanine-N2 methyltransferase. *RNA N Y N* **15**: 666–674.

Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*.

Bilsland E, Sparkes A, Williams K, Moss HJ, Clare M de, Pir P, Rowland J, Aubrey W, Pateman R, Young M, et al. 2013. Yeast-based automated high-throughput screens to identify anti-parasitic lead compounds. *Open Biol* **3**: 120158.

Black CB, Duensing TD, Trinkle LS, Dunlay RT. 2011. Cell-based screening using high-throughput flow cytometry. *Assay Drug Dev Technol* **9**: 13–20.

Bockus AT, McEwen CM, Lokey RS. 2013. Form and function in cyclic peptide natural products: a pharmacokinetic perspective. *Curr Top Med Chem* **13**: 821–836.

Boelz S, Neu-Yilik G, Gehring NH, Hentze MW, Kulozik AE. 2006. A chemiluminescence-based reporter system to monitor nonsense-mediated mRNA decay. *Biochem Biophys Res Commun* **349**: 186–191.

Bono F, Gehring NH. 2011. Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biol* **8**: 24–29.

Boon K-L, Kos M. 2010. Deletion of Swm2p selectively impairs trimethylation of snRNAs by trimethylguanosine synthase (Tgs1p). *FEBS Lett* **584**: 3299–3304.

Brewster NK, Val DL, Walker ME, Wallace JC. 1994. Regulation of Pyruvate Carboxylase Isozyme (PYC1, PYC2) Gene Expression in Saccharomyces cerevisiae during Fermentative and Nonfermentative Growth. *Arch Biochem Biophys* **311**: 62–71.

Brodsky AS, Silver PA. 2000. Pre-mRNA processing factors are required for nuclear export. *RNA* **6**: 1737–1749.

Brugiolo M, Herzel L, Neugebauer KM. 2013. Counting on co-transcriptional splicing. *F1000Prime Rep* **5**.

Buschini A, Poli P, Rossi C. 2003. Saccharomyces cerevisiae as an eukaryotic cell model to assess cytotoxicity and genotoxicity of three anticancer anthraquinones. *Mutagenesis* **18**: 25–36.

Cadwell RC, Joyce GF. 1992. Randomization of genes by PCR mutagenesis. *Genome Res* **2**: 28–33.

Caporilli S, Yu Y, Jiang J, White-Cooper H. 2013. The RNA Export Factor, Nxt1, Is Required for Tissue Specific Transcriptional Regulation. *PLoS Genet* **9**: e1003526.

Chan S, Choi E-A, Shi Y. 2011. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2**: 321–335.

Chang J, Schwer B, Shuman S. 2010. Mutational analyses of trimethylguanosine synthase (Tgs1) and Mud2: proteins implicated in pre-mRNA splicing. *RNA* **16**: 1018–1031.

Chávez S, Beilharz T, Rondón AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *EMBO J* **19**: 5824–5834.

Chen J, Young SM, Allen C, Seeber A, Péli-Gulli M-P, Panchaud N, Waller A, Ursu O, Yao T, Golden JE, et al. 2012. Identification of a small molecule yeast TORC1 inhibitor with a multiplex screen based on flow cytometry. *ACS Chem Biol* **7**: 715–722.

Chen Q-H. 2012. Review of Kinase Inhibitor Drugs. *J Nat Prod* **75**: 2269–2269.

Chen Y-H, Su L-H, Sun C-H. 2008. Incomplete nonsense-mediated mRNA decay in Giardia lamblia. *Int J Parasitol* **38**: 1305–1317.

Chernikova SB, Razorenova OV, Higgins JP, Sishc BJ, Nicolau M, Dorth JA, Chernikova DA, Kwok S, Brooks JD, Bailey SM, et al. 2012. Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability. *Cancer Res* **72**: 2111–2119.

Chlebowski A, Lubas M, Jensen TH, Dziembowski A. 2013. RNA decay machines: The exosome. *Biochim Biophys Acta BBA - Gene Regul Mech* **1829**: 552–560.

Colau G, Thiry M, Leduc V, Bordonné R, Lafontaine DLJ. 2004. The small nucle(ol)ar RNA cap trimethyltransferase is required for ribosome synthesis and intact nucleolar morphology. *Mol Cell Biol* **24**: 7976–7986.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* **136**: 777–793.

Corbett A, Pickett J, Burns A, Corcoran J, Dunnett SB, Edison P, Hagan JJ, Holmes C, Jones E, Katona C, et al. 2012. Drug repositioning for Alzheimer's disease. *Nat Rev Drug Discov* **11**: 833–846.

Cougot N, van Dijk E, Babajko S, Séraphin B. 2004. "Cap-tabolism." *Trends Biochem Sci* **29**: 436–444.

Cragg GM, Grothaus PG, Newman DJ. 2014. New horizons for old drugs and drug leads. *J Nat Prod* **77**: 703–723.

Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**: 292–301.

Crick FHC. 1970. Central Dogma of Molecular Biology. *Nature* **227**: 561–563.

Crick FHC. 1958. On protein synthesis. In *Symp Soc Exp Bio*, Vol. XII of, pp. 139–163.

Crook JM, Kobayashi NR. 2008. Human stem cells for modeling neurological disorders: accelerating the drug discovery pipeline. *J Cell Biochem* **105**: 1361–1366.

Das B, Guo Z, Russo P, Chartrand P, Sherman F. 2000. The role of nuclear cap binding protein Cbc1p of yeast in mRNA termination and degradation. *Mol Cell Biol* **20**: 2827–2838.

Decottignies A, Grant AM, Nichols JW, de Wet H, McIntosh DB, Goffeau A. 1998. ATPase and multidrug transport activities of the overexpressed yeast ABC protein Yor1p. *J Biol Chem* **273**: 12612–12622.

Dehm SM. 2013. Test-firing ammunition for spliceosome inhibition in cancer. *Clin Cancer Res Off J Am Assoc Cancer Res* **19**: 6064–6066.

Delaveau T, Delahodde A, Carvajal E, Subik J, Jacq C. 1994. PDR3, a new yeast regulatory gene, is homologous to PDR1 and controls the multidrug resistance phenomenon. *Mol Gen Genet MGG* **244**: 501–511.

Dieppois G, Iglesias N, Stutz F. 2006. Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes. *Mol Cell Biol* **26**: 7858–7870.

DiMasi JA, Hansen RW, Grabowski HG. 2003. The price of innovation: new estimates of drug development costs. *J Health Econ* **22**: 151–185.

Dower K, Kuperwasser N, Merrikh H, Rosbash M. 2004. A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. *RNA* **10**: 1888–1899.

Dower K, Rosbash M. 2002. T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA* **8**: 686–697.

Dreumont N, Séraphin B. 2013. Rapid screening of yeast mutants with reporters identifies new splicing phenotypes. *FEBS J* **280**: 2712–2726.

Drews J. 2000. Drug Discovery: A Historical Perspective. *Science* **287**: 1960–1964.

Driggers EM, Hale SP, Lee J, Terrett NK. 2008. The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat Rev Drug Discov* **7**: 608–624.

Dufu K, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, Reed R. 2010. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes Dev* **24**: 2043–2053.

Dunckley T, Parker R. 1999. The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *EMBO J* **18**: 5411–5422.

Dziembowski A, Lorentzen E, Conti E, Séraphin B. 2007. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol* **14**: 15–22.

Edery I, Sonenberg N. 1985. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proc Natl Acad Sci U S A* **82**: 7590–7594.

Edwards BS, Kuckuck FW, Prossnitz ER, Ransom JT, Sklar LA. 2001. HTPS flow cytometry: a novel platform for automated high throughput drug discovery and characterization. *J Biomol Screen* **6**: 83–90.

Edwards BS, Young SM, Saunders MJ, Bologa C, Oprea TI, Ye RD, Prossnitz ER, Graves SW, Sklar LA. 2007. High-throughput flow cytometry for drug discovery. *Expert Opin Drug Discov* **2**: 685–696.

Effenberger KA, Anderson DD, Bray WM, Prichard BE, Ma N, Adams MS, Ghosh AK, Jurica MS. 2014. Coherence between cellular responses and in vitro splicing inhibition for the anti-tumor drug pladienolide B and its analogs. *J Biol Chem* **289**: 1938–1947.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U A* **95**: 14863–8.

Ellisdon AM, Jani D, Kohler A, Hurt E, Stewart M. 2010. Structural basis for the interaction between yeast Spt-Ada-Gcn5 acetyltransferase (SAGA) complex components Sgf11 and Sus1. *J Biol Chem* **285**: 3850–6.

Engle SJ, Vincent F. 2014. Small molecule screening in human induced pluripotent stem cell-derived terminal cell types. *J Biol Chem* **289**: 4562–4570.

Fabbro D, Cowan-Jacob SW, Möbitz H, Martiny-Baron G. 2012. Targeting cancer with small-molecular-weight kinase inhibitors. *Methods Mol Biol Clifton NJ* **795**: 1–34.

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Lührmann R. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36**: 593–608.

Fang F, Hoskins J, Butler JS. 2004. 5-fluorouracil enhances exosome-dependent accumulation of polyadenylated rRNAs. *Mol Cell Biol* **24**: 10766–10776.

Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and disease. *Genes Dev* **17**: 419–437.

Fillingham J, Kainth P, Lambert J-P, van Bakel H, Tsui K, Peña-Castillo L, Nislow C, Figeys D, Hughes TR, Greenblatt J, et al. 2009. Two-Color Cell Array Screen Reveals Interdependent Roles for Histone Chaperones and a Chromatin Boundary Regulator in Histone Gene Repression. *Mol Cell* **35**: 340–351.

Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E. 2002. The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *EMBO J* **21**: 5843–52.

Fischer U, Englbrecht C, Chari A. 2011. Biogenesis of spliceosomal small nuclear ribonucleoproteins. *Wiley Interdiscip Rev RNA* **2**: 718–731.

Fischer U, Sumpter V, Sekine M, Satoh T, Lührmann R. 1993. Nucleocytoplasmic Transport of U snRNPs - Definition of a nuclear localization signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J* **12**: 573–583.

Fortney K, Xie W, Kotlyar M, Griesman J, Kotseruba Y, Jurisica I. 2013. NetwoRx: connecting drugs to networks and phenotypes in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **41**: D720–727.

Fouser LA, Friesen JD. 1986. Mutations in a yeast intron demonstrate the importance of specific conserved nucleotides for the two stages of nuclear mRNA splicing. *Cell* **45**: 81–93.

Frank SA. 2013. Evolution of robustness and cellular stochasticity of gene expression. *PLoS Biol* **11**: e1001578.

Franke J, Gehlen J, Ehrenhofer-Murray AE. 2008. Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J Cell Sci* **121**: 3553–3560.

Fresco LD, Buratowski S. 1996. Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA* **2**: 584–596.

Fujimura S, Xing Y, Takeya M, Yamashita Y, Ohshima K, Kuwahara K, Sakaguchi N. 2005. Increased expression of germinal center-associated nuclear protein RNA-primase is associated with lymphomagenesis. *Cancer Res* **65**: 5925–5934.

Gaber RF, Copple DM, Kennedy BK, Vidal M, Bard M. 1989. The yeast gene ERG6 is required for normal membrane function but is not essential for biosynthesis of the cell-cycle-sparking sterol. *Mol Cell Biol* **9**: 3447–3456.

Gadal O, Strauss D, Kessl J, Trumpower B, Tollervey D, Hurt E. 2001. Nuclear export of 60S ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p that associates with the large subunit protein Rpl10p. *Mol Cell Biol* **21**: 3405–3415.

Gallie DR. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* **5**: 2108–2116.

Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A, Nehrbass U. 2004. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116**: 63–73.

Geitz RD, Scheistel RH. 1995. Transforming yeast with DNA. *Methods Mol Cell Biol* **5**: 255–269.

Ghosh A, Lima CD. 2010. Enzymology of RNA cap synthesis. *Wiley Interdiscip Rev RNA* **1**: 152–172.

Ghosh I, Sun L, Xu M-Q. 2001. Zinc Inhibition of Protein trans-Splicing and Identification of Regions Essential for Splicing and Association of a Split Intein*. *J Biol Chem* **276**: 24051–24058.

Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW. 2004. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci U S A* **101**: 793–798.

Gietz RD, Schiestl RH. 2007a. Microtiter plate transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**: 5–8.

Gietz RD, Schiestl RH. 2007b. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**: 35–37.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. In *Guide to Yeast Genetics and Molecular and Cell Biology, Pt B*, Vol. 350 of *Methods in Enzymology*, pp. 87–96 ://000176466300004.

Goldstrohm AC, Wickens M. 2008. Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* **9**: 337–44.

González CI, Bhattacharya A, Wang W, Peltz SW. 2001. Nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Gene* **274**: 15–25.

Green DM, Johnson CP, Hagan H, Corbett AH. 2003. The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proc Natl Acad Sci U S A* **100**: 1010–1015.

Griffith EC, Licitra EJ, Liu JO. 2000. Yeast three-hybrid system for detecting ligand-receptor interactions. *Methods Enzymol* **328**: 89–103.

Grosjean H. 2005. Modification and editing of RNA: historical overview and important facts to remember. In *Fine-Tuning of RNA Functions by Modification and Editing* (ed. H. Grosjean), *Topics in Current Genetics*, pp. 1–22, Springer Berlin Heidelberg

Grosjean H, Benne R. 1998. *Modification and Editing of Rna*. ASM Press.

Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JEG, Wagner G. 2003. Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E. *Cell* **115**: 739–750.

Grosso AR, Martins S, Carmo-Fonseca M. 2008. The emerging role of splicing factors in cancer. *EMBO Rep* **9**: 1087–1093.

Gruter P, Tabernero C, von Kobbe C, Schmitt C, Saavedra C, Bachi A, Wilm M, Felber BK, Izaurralde E. 1998. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol Cell* **1**: 649–659.

Gu J, Patton JR, Shimba S, Reddy R. 1996. Localization of modified nucleotides in Schizosaccharomyces pombe spliceosomal small nuclear RNAs: modified nucleotides are clustered in functionally important regions. *RNA* **2**: 909–918.

Guarente L, Ptashne M. 1981. Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **78**: 2199–2203.

Guarente L, Yocum RR, Gifford P. 1982. A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. *Proc Natl Acad Sci* **79**: 7410–7414.

Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M. 2004. A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* **32**: 5379–91.

Gupta SC, Sung B, Prasad S, Webb LJ, Aggarwal BB. 2013. Cancer drug discovery by repurposing: teaching new tricks to old dogs. *Trends Pharmacol Sci* **34**: 508–517.

Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, Stutz F, Dargemont C. 2006. Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proc Natl Acad Sci U S A* **103**: 16376–16381.

Hackmann A, Wu H, Schneider U-M, Meyer K, Jung K, Krebber H. 2014. Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat Commun* **5**.

Hadjiargyrou M, Delihas N. 2013. The Intertwining of Transposable Elements and Non-Coding RNAs. *Int J Mol Sci* **14**: 13307–13328.

Hagan KW, Ruiz-Echevarria MJ, Quan Y, Peltz SW. 1995. Characterization of cis-acting sequences and decay intermediates involved in nonsense-mediated mRNA turnover. *Mol Cell Biol* **15**: 809–823.

Haimovich G, Medina DA, Causse SZ, Garber M, Millán-Zambrano G, Barkai O, Chávez S, Pérez-Ortín JE, Darzacq X, Choder M. 2013. Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* **153**: 1000–1011.

Hamm J, Darzynkiewicz E, Tahara SM, Mattaj IW. 1990. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* **62**: 569–577.

Hammell CM, Gross S, Zenklusen D, Heath CV, Stutz F, Moore C, Cole CN. 2002. Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* **22**: 6441–6457.

Harigaya Y, Parker R. 2010. No-go decay: a quality control mechanism for RNA in translation. *Wiley Interdiscip Rev RNA* **1**: 132–141.

Harrison C. 2012. Kinase inhibitors: Analysing kinase inhibitor selectivity. *Nat Rev Drug Discov* **11**: 21–21.

Hartwell L. 1967. Macromolecule synthesis in temperature-sensitive mutants of yeast. *J Bact* **93**: 1662–1670.

Hasslacher M, Ivessa AS, Paltauf F, Kohlwein SD. 1993. Acetyl-CoA carboxylase from yeast is an essential enzyme and is regulated by factors that control phospholipid metabolism. *J Biol Chem* **268**: 10946–10952.

Hausmann S, Ramirez A, Schneider S, Schwer B, Shuman S. 2007. Biochemical and genetic analysis of RNA cap guanine-N2 methyltransferases from Giardia lamblia and Schizosaccharomyces pombe. *Nucleic Acids Res* **35**: 1411–1420.

Hausmann S, Shuman S. 2005. Specificity and Mechanism of RNA Cap Guanine-N2 Methyltransferase (Tgs1). *J Biol Chem* **280**: 4021–4024.

Hausmann S, Zheng S, Costanzo M, Brost RL, Garcin D, Boone C, Shuman S, Schwer B. 2008. Genetic and Biochemical Analysis of Yeast and Human Cap Trimethylguanosine Synthase. *J Biol Chem* **283**: 31706–31718.

He F, Brown AH, Jacobson A. 1997. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol Cell Biol* **17**: 1580–94.

He F, Jacobson A. 1995. Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes Dev* **9**: 437–454.

He F, Peltz SW, Donahue JL, Rosbash M, Jacobson A. 1993. Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upfl- mutant. *Proc Natl Acad Sci USA* **90**: 7034–7038.

Heim R, Cubitt AB, Tsien RY. 1995. Improved green fluorescence. *Nature* **373**: 663–4.

Heitman J, Movva NR, Hall MN. 1991. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **253**: 905–909.

Henthorn DC, Jaxa-Chamiec AA, Meldrum E. 2002. A GAL4-based yeast three-hybrid system for the identification of small molecule-target protein interactions. *Biochem Pharmacol* **63**: 1619–1628.

Hernandez N. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* **276**: 26733–26736.

Herold A, Suyama M, Rodrigues JP, Braun IC, Kutay U, Carmo-Fonseca M, Bork P, Izaurralde E. 2000. TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Mol Cell Biol* **20**: 8996–9008.

Herrmann H. 2013. Nuclear architecture and dynamics: chromatin, epigenetics, genomics: Review of Genome Organization and Function in the Cell Nucleus: Edited by Karsten Rippe. *Nucleus* **4**: 85–88.

Hessle V, von Euler A, González de Valdivia E, Visa N. 2012. Rrp6 is recruited to transcribed genes and accompanies the spliced mRNA to the nuclear pore. *RNA N Y N* **18**: 1466–1474.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, et al. 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**: 362–365.

Hilleren P, Parker R. 2001. Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3 ’-end formation of nascent transcripts. *RNA* **7**: 753–764.

Hoja U, Marthol S, Hofmann J, Stegner S, Schulz R, Meier S, Greiner E, Schweizer E. 2004. HFA1 Encoding an Organelle-specific Acetyl-CoA Carboxylase Controls Mitochondrial Fatty Acid Synthesis in Saccharomyces cerevisiae. *J Biol Chem* **279**: 21779–21786.

Hollstein U. 1974. Actinomycin. Chemistry and mechanism of action. *Chem Rev* **74**: 625–652.

Hoose SA, Duran C, Malik I, Eslamfam S, Shasserre SC, Downing SS, Hoover EM, Dowd KE, Smith R 3rd, Polymenis M. 2012. Systematic analysis of cell cycle effects of common drugs leads to the discovery of a suppressive interaction between gemfibrozil and fluoxetine. *PloS One* **7**: e36503.

Horswill AR, Benkovic SJ. 2005. Cyclic peptides, a chemical genetics tool for biologists. *Cell Cycle Georget Tex* **4**: 552–555.

Horswill AR, Savinov SN, Benkovic SJ. 2004. A systematic method for identifying small-molecule modulators of protein-protein interactions. *Proc Natl Acad Sci U S A* **101**: 15591–15596.

Houseley J, LaCava J, Tollervey D. 2006. RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* **7**: 529–539.

Houtman JCD, Houghtling RA, Barda-Saad M, Toda Y, Samelson LE. 2005. Early phosphorylation kinetics of proteins involved in proximal TCR-mediated signaling pathways. *J Immunol Baltim Md 1950* **175**: 2449–2458.

Howson R, Huh W-K, Ghaemmaghami S, Falvo JV, Bower K, Belle A, Dephoure N, Wykoff DD, Weissman JS, O'Shea EK. 2005. Construction, verification and experimental use of two epitope-tagged collections of budding yeast strains. *Comp Funct Genomics* **6**: 2–16.

Huang P, Sandoval A, Van Den Neste E, Keating MJ, Plunkett W. 2000. Inhibition of RNA transcription: a biochemical mechanism of action against chronic lymphocytic leukemia cells by fludarabine. *Leukemia* **14**: 1405–1413.

Huang Y, Carmichael GG. 1996. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol* **16**: 1534–1542.

Huber J, Cronshagen U, Kadokura M, Marshallsay C, Wada T, Sekine M, Lührmann R. 1998. Snurportin1, an m(3)G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J* **17**: 4114–4126.

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. *Nature* **425**: 686–691.

Hung N-J, Lo K-Y, Patel SS, Helmke K, Johnson AW. 2008. Arx1 is a nuclear export receptor for the 60S ribosomal subunit in yeast. *Mol Biol Cell* **19**: 735–744.

Hurt JA, Silver PA. 2008. mRNA nuclear export and human disease. *Dis Model Mech* **1**: 103–108.

Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Von Dach E, Corbett AH, Dargemont C, Stutz F. 2010. Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes Dev* **24**: 1927–1938.

Inoue H, Nagata N, Kurokawa H, Yamanaka S. 2014. iPS cells: a game changer for future medicine. *EMBO J* **33**: 409–417.

Iorio MV, Croce CM. 2012. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol Med* **4**: 143–159.

Irish JM, Hovland R, Krutzik PO, Perez OD, Bruserud Ø, Gjertsen BT, Nolan GP. 2004. Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* **118**: 217–228.

Jacquier A, Rodriguez JR, Rosbash M. 1985. A quantitative analysis of the effects of 5′ junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. *Cell* **43**: 423–430.

Jády BE, Bertrand E, Kiss T. 2004. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body–specific localization signal. *J Cell Biol* **164**: 647–652.

James SA, Turner W, Schwer B. 2002. How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**: 1068–1077.

Jani D, Lutz S, Marshall NJ, Fischer T, Köhler A, Ellisdon AM, Hurt E, Stewart M. 2009. Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. *Mol Cell* **33**: 727–737.

Jensen J, Hyllner J, Björquist P. 2009. Human embryonic stem cell technologies and drug discovery. *J Cell Physiol* **219**: 513–519.

Jin G, Wong STC. 2014. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov Today* **19**: 637–644.

Johnson BA, Anker H, Meleney FL. 1945. BACITRACIN: A NEW ANTIBIOTIC PRODUCED BY A MEMBER OF THE B. SUBTILIS GROUP. *Science* **102**: 376–377.

Johnston IG. 2012. The chaos within: exploring noise in cellular biology. *Significance* **9**: 17–21.

Jones AW. 2011. Early drug discovery and the rise of pharmaceutical chemistry. *Drug Test Anal* **3**: 337–344.

Joo SH. 2012. Cyclic Peptides as Therapeutic Agents and Biochemical Tools. *Biomol Ther* **20**: 19–26.

Juneau K, Miranda M, Hillenmeyer ME, Nislow C, Davis RW. 2006. Introns Regulate RNA and Protein Abundance in Yeast. *Genetics* **174**: 511–518.

Juneau K, Palm C, Miranda M, Davis RW. 2007. High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci U S A* **104**: 1522–1527.

Kageshita T, Kuwahara K, Oka M, Ma D, Ono T, Sakaguchi N. 2006. Increased expression of germinal center-associated nuclear protein (GANP) is associated with malignant transformation of melanocytes. *J Dermatol Sci* **42**: 55–63.

Kahan BD. 1985. Cyclosporine: the agent and its actions. *Transplant Proc* **17**: 5–18.

Kainth P, Andrews B. 2010. Illuminating transcription pathways using fluorescent reporter genes and yeast functional genomics. *Transcription* **1**: 76–80.

Kainth P, Sassi HE, Peña-Castillo L, Chua G, Hughes TR, Andrews B. 2009. Comprehensive genetic analysis of transcription factor pathways using a dual reporter gene system in budding yeast. *Methods* **48**: 258–264.

Kalina T, Stuchlý J, Janda A, Hrusák O, Růzicková S, Sedivá A, Litzman J, Vlková M. 2009. Profiling of polychromatic flow cytometry data on B-cells reveals patients' clusters in common variable immunodeficiency. *Cytom Part J Int Soc Anal Cytol* **75**: 902–909.

Kambach C, Walke S, Nagai K. 1999a. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol* **9**: 222–230.

Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, Nagai K. 1999b. Crystal structure of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**: 375–387.

Kammler S, Lykke-Andersen S, Jensen TH. 2008. The RNA Exosome Component hRrp6 Is a Target for 5-Fluorouracil in Human Cells. *Mol Cancer Res* **6**: 990–995.

Kang YB, Cullen BR. 1999. The human Tap protein is a nuclear mRNA export factor that contains novel RNA-binding and nucleocytoplasmic transport sequences. *Genes Dev* **13**: 1126–1139.

Karijolich J, Yu Y-T. 2010. Spliceosomal snRNA modifications and their function. *RNA Biol* **7**: 192–204.

Karin M, Najarian R, Haslinger A, Valenzuela P, Welch J, Fogel S. 1984. Primary structure and transcription of an amplified genetic locus: the CUP1 locus of yeast. *Proc Natl Acad Sci U S A* **81**: 337–341.

Kawakami T, Ohta A, Ohuchi M, Ashigai H, Murakami H, Suga H. 2009. Diverse backbone-cyclized peptides via codon reprogramming. *Nat Chem Biol* **5**: 888–890.

Kim HS, Hoja U, Stolz J, Sauer G, Schweizer E. 2004. Identification of the tRNA-binding Protein Arc1p as a Novel Target of in Vivo Biotinylation in Saccharomyces cerevisiae. *J Biol Chem* **279**: 42445–42452.

Kimura RH, Tran A-T, Camarero JA. 2006. Biosynthesis of the cyclotide Kalata B1 by using protein splicing. *Angew Chem Int Ed Engl* **45**: 973–976.

Kinsella TM, Ohashi CT, Harder AG, Yam GC, Li W, Peelle B, Pali ES, Bennett MK, Molineaux SM, Anderson DA, et al. 2002. Retrovirally delivered random cyclic Peptide libraries yield inhibitors of interleukin-4 signaling in human B cells. *J Biol Chem* **277**: 37512–37518.

Kiss T. 2004. Biogenesis of small nuclear RNPs. *J Cell Sci* **117**: 5949–5951.

Kitsera N, Khobta A, Epe B. 2007. Destabilized green fluorescent protein detects rapid removal of transcription blocks after genotoxic exposure. *BioTechniques* **43**: 222–227.

Komili S, Silver PA. 2008. Coupling and coordination in gene expression processes: a systems biology view. *Nat Rev Genet* **9**: 38–48.

Komonyi O, Pápai G, Enunlu I, Muratoglu S, Pankotai T, Kopitova D, Maróy P, Udvardy A, Boros I. 2005. DTL, the Drosophila homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development. *J Biol Chem* **280**: 12397–12404.

Konarska MM, Padgett RA, Sharp PA. 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **38**: 731–736.

Kong K-YE, Tang H-MV, Pan K, Huang Z, Lee T-HJ, Hinnebusch AG, Jin D-Y, Wong C-M. 2014. Cotranscriptional recruitment of yeast TRAMP complex to intronic sequences promotes optimal pre-mRNA splicing. *Nucleic Acids Res* **42**: 643–660.

Krainer AR, Maniatis T, Ruskin B, Green MR. 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* **36**: 993–1005.

Kress TL, Krogan NJ, Guthrie C. 2008. A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol Cell* **32**: 727–734.

Kritzer JA, Hamamichi S, McCaffery JM, Santagata S, Naumann TA, Caldwell KA, Caldwell GA, Lindquist S. 2009. Rapid selection of cyclic peptides that reduce alpha-synuclein toxicity in yeast and animal models. *Nat Chem Biol* **5**: 655–663.

Kulak O, Lum L. 2013. A Multiplexed Luciferase-based Screening Platform for Interrogating Cancer-associated Signal Transduction in Cultured Cells. *J Vis Exp JoVE*.

Kurtovic-Kozaric A, Przychodzen B, Singh J, Konarska MM, Clemente MJ, Otrock ZK, Nakashima M, Hsi ED, Yoshida K, Ogawa S, et al. 2014. PRPF8 defects cause missplicing in myeloid malignancies. *Leukemia*.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**: 713–724.

LaGrandeur TE, Parker R. 1998. Isolation and characterization of Dcp1p, the yeast mRNA decapping enzyme. *EMBO J* **17**: 1487–1496.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Landrigan A, Wong MT, Utz PJ. 2011. CpG and non-CpG oligodeoxynucleotides directly costimulate mouse and human CD4+ T cells through a TLR9- and MyD88-independent mechanism. *J Immunol Baltim Md 1950* **187**: 3033–3043.

Larimer FW, Hsu CL, Maupin MK, Stevens A. 1992a. Characterization of the XRN1 gene encoding a 5'-->3' exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells. *Gene* **120**: 51–57.

Larimer FW, Hsu CL, Maupin MK, Stevens A. 1992b. Characterization Of The Xrn1 Gene Encoding A 5'->3' Exoribonuclease - Sequence Data And Analysis Of Disparate Protein And Messenger-RNA Levels Of Gene-Disrupted Yeast-Cells. *Gene* **120**: 51–57.

Larimer FW, Stevens A. 1990. Disruption Of The Gene Xrn1, Coding For A 5'->3' Exoribonuclease, Restricts Yeast-Cell Growth. *Gene* **95**: 85–90.

Larson GP, Itakura K, Ito H, Rossi JJ. 1983. Saccharomyces cerevisiae actin--Escherichia coli lacZ gene fusions: synthetic-oligonucleotide-mediated deletion of the 309 base pair intervening sequence in the actin gene. *Gene* **22**: 31–39.

Leeds P, Peltz SW, Jacobson A, Culbertson MR. 1991. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* **5**: 2303–2314.

Legrain P, Rosbash M. 1989. Some cis-acting and trans-acting mutants for splicing target pre-messenger RNA to the cytoplasm. *Cell* **57**: 573–583.

Legrain P, Seraphin B, Rosbash M. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol Cell Biol* **8**: 3755–3760.

Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bähler J, Bachand F. 2011. A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol Cell* **44**: 108–119.

Lesser CF, Guthrie C. 1993a. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **133**: 851–63.

Lesser CF, Guthrie C. 1993b. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**: 1982–8.

Li D-Q, Pakala SB, Nair SS, Eswaran J, Kumar R. 2012. Metastasis-associated protein 1/nucleosome remodeling and histone deacetylase complex in cancer. *Cancer Res* **72**: 387–394.

Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang C-C, Kain SR. 1998. Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. *J Biol Chem* **273**: 34970–34975.

Liao XC, Tang J, Rosbash M. 1993. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. *Genes Dev* **7**: 419–428.

Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH. 2002. Interactions between mRNA export commitment, 3 ’-end quality control, and nuclear degradation. *Mol Cell Biol* **22**: 8254–8266.

Locher KP. 2009. Review. Structure and mechanism of ATP-binding cassette transporters. *Philos Trans R Soc Lond B Biol Sci* **364**: 239–245.

Longley DB, Harkin DP, Johnston PG. 2003. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer* **3**: 330–338.

Longtine MS, McKenzie A, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**: 953–961.

Lum PY, Armour CD, Stepaniants SB, Cavet G, Wolf MK, Butler JS, Hinshaw JC, Garnier P, Prestwich GD, Leonardson A, et al. 2004. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell* **116**: 121–137.

Luna R, Rondón AG, Aguilera A. 2012. New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. *Biochim Biophys Acta* **1819**: 514–520.

Lundblad V, Struhl K. 2008. Chapter 13: Yeast. In *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.

Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, Green DVS, Hertzberg RP, Janzen WP, Paslay JW, et al. 2011. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **10**: 188–195.

Malcovati L, Papaemmanuil E, Bowen DT, Boultwood J, Della Porta MG, Pascutto C, Travaglino E, Groves MJ, Godfrey AL, Ambaglio I, et al. 2011. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* **118**: 6239–6246.

Mamnun YM, Pandjaitan R, Mahé Y, Delahodde A, Kuchler K. 2002. The yeast zinc finger regulators Pdr1p and Pdr3p control pleiotropic drug resistance (PDR) as homo- and heterodimers in vivo. *Mol Microbiol* **46**: 1429–1440.

Mandel HG. 1981. The target cell determinants of the antitumor actions of 5-FU: does FU incorporation into RNA play a role? *Cancer Treat Rep* **65 Suppl 3**: 63–71.

Marx UC, Korsinczky MLJ, Schirra HJ, Jones A, Condie B, Otvos L Jr, Craik DJ. 2003. Enzymatic cyclization of a potent bowman-birk protease inhibitor, sunflower trypsin inhibitor-1, and solution structure of an acyclic precursor peptide. *J Biol Chem* **278**: 21782–21789.

Massenet S, Motorin Y, Lafontaine DLJ, Hurt EC, Grosjean H, Branlant C. 1999. Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol* **19**: 2142–2154.

Mattaj IW. 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**: 905–911.

Mattaj IW, De Robertis EM. 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell* **40**: 111–118.

Mayas RM, Maita H, Staley JP. 2006. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* **13**: 482–490.

Metzstein MM, Krasnow MA. 2006. Functions of the Nonsense-Mediated mRNA Decay Pathway in Drosophila Development. *PLoS Genet* **2**.

Misteli T. 2005. Concepts in nuclear architecture. *BioEssays News Rev Mol Cell Dev Biol* **27**: 477–487.

Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**: 343–8.

Moehle EA, Ryan CJ, Krogan NJ, Kress TL, Guthrie C. 2012. The yeast SR-like protein Npl3 links chromatin modification to mRNA processing. *PLoS Genet* **8**: e1003101.

Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. 2011. The cost of drug development: a systematic review. *Health Policy Amst Neth* **100**: 4–17.

Mouaikel J, Bujnicki JM, Tazi J, Bordonné R. 2003. Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res* **31**: 4899–4909.

Mouaikel J, Verheggen C, Bertrand E, Tazi J, Bordonné R. 2002. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell* **9**: 891–901.

Muhlrad D, Decker CJ, Parker R. 1994. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes Dev* **8**: 855–866.

Muhlrad D, Decker CJ, Parker R. 1995. Turnover mechanisms of the stable yeast PGK1 mRNA. *Mol Cell Biol* **15**: 2145–2156.

Muhlrad D, Parker R. 1999a. Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **5**: 1299–307.

Muhlrad D, Parker R. 1999b. Recognition of yeast mRNAs as "nonsense containing" leads to both inhibition of mlRNA translation and mRNA degradation: Implications for the control of mRNA decapping. *Mol Biol Cell* **10**: 3971–3978.

Narayanan U, Achsel T, Lührmann R, Matera AG. 2004. Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol Cell* **16**: 223–234.

Narayanan U, Ospina JK, Frey MR, Hebert MD, Matera AG. 2002. SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Hum Mol Genet* **11**: 1785–1795.

Nilsen TW. 2003. The spliceosome: the most complex macromolecular machine in the cell? *BioEssays* **25**: 1147–1149.

Noble SM, Guthrie C. 1996. Transcriptional pulse-chase analysis reveals a role for a novel snRNP-associated protein in the manufacture of spliceosomal snRNPs. *EMBO J* **15**: 4368–4379.

Nourani A, Wesolowski-Louvel M, Delaveau T, Jacq C, Delahodde A. 1997. Multiple-drug-resistance phenomenon in the yeast Saccharomyces cerevisiae: involvement of two hexose transporters. *Mol Cell Biol* **17**: 5453–5460.

Oesterreich FC, Bieberstein N, Neugebauer KM. 2011. Pause locally, splice globally. *Trends Cell Biol* **21**: 328–35.

Oldham ML, Davidson AL, Chen J. 2008. Structural insights into ABC transporter mechanism. *Curr Opin Struct Biol* **18**: 726–733.

Oltean S, Bates DO. 2013. Hallmarks of alternative splicing in cancer. *Oncogene*.

Ong S-E, Li X, Schenone M, Schreiber SL, Carr SA. 2012. Identifying cellular targets of small-molecule probes and drugs with biochemical enrichment and SILAC. *Methods Mol Biol Clifton NJ* **803**: 129–140.

Ong S-E, Schenone M, Margolin AA, Li X, Do K, Doud MK, Mani DR, Kuai L, Wang X, Wood JL, et al. 2009. Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc Natl Acad Sci U S A* **106**: 4617–4622.

Paillusson A, Hirschi N, Vallan C, Azzalin CM, Muhlemann O. 2005. A GFP-based reporter system to monitor nonsense-mediated mRNA decay. *Nucleic Acids Res* **33**: e54.

Palacios I, Hetzer M, Adam SA, Mattaj IW. 1997. Nuclear import of U snRNPs requires importin beta. *EMBO J* **16**: 6783–6792.

Palancade B, Zuccolo M, Loeillet S, Nicolas A, Doye V. 2005. Pml39, a novel protein of the nuclear periphery required for nuclear retention of improper messenger ribonucleoparticles. *Mol Biol Cell* **16**: 5258–5268.

Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C, et al. 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* **365**: 1384–1395.

Parenteau J, Durand M, Véronneau S, Lacombe A-A, Morin G, Guérin V, Cecez B, Gervais-Bird J, Koh C-S, Brunelle D, et al. 2008. Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol Biol Cell* **19**: 1932–1941.

Parker R, Guthrie C. 1985. A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation. *Cell* **41**: 107–118.

Parker R, Sheth U. 2007. P Bodies and the Control of mRNA Translation and Degradation. *Mol Cell* **25**: 635–646.

Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, Brown GW, Kane PM, Hughes TR, Boone C. 2004. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol* **22**: 62–69.

Parsons AB, Lopez A, Givoni IE, Williams DE, Gray CA, Porter J, Chua G, Sopko R, Brost RL, Ho C-H, et al. 2006. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* **126**: 611–625.

Patel SB, Bellini M. 2008. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**: 6482–6493.

Patton JR, Padgett RW. 2005. Pseudouridine modification in Caenorhabditis elegans spliceosomal snRNAs: unique modifications are found in regions involved in snRNA-snRNA interactions. *BMC Mol Biol* **6**: 20.

Paul S, Moye-Rowley WS. 2014. Multidrug resistance in fungi: regulation of transporter-encoding gene expression. *Front Physiol* **5**: 143.

Peña MM, Koch KA, Thiele DJ. 1998. Dynamic regulation of copper uptake and detoxification genes in Saccharomyces cerevisiae. *Mol Cell Biol* **18**: 2514–2523.

Pereira C, Leão M, Soares J, Bessa C, Saraiva L. 2012. New therapeutic strategies for cancer and neurodegeneration emerging from yeast cell-based systems. *Curr Pharm Des* **18**: 4223–4235.

Perkins E, Sun D, Nguyen A, Tulac S, Francesco M, Tavana H, Nguyen H, Tugendreich S, Barthmaier P, Couto J, et al. 2001. Novel inhibitors of poly(ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. *Cancer Res* **61**: 4175–4183.

Petrausch U, Haley D, Miller W, Floyd K, Urba WJ, Walker E. 2006. Polychromatic flow cytometry: a rapid method for the reduction and analysis of complex multiparameter data. *Cytom Part J Int Soc Anal Cytol* **69**: 1162–1173.

Phillips BW, Crook JM. 2010. Pluripotent human stem cells: A novel tool in drug discovery. *BioDrugs Clin Immunother Biopharm Gene Ther* **24**: 99–108.

Phimsen S, Kuwahara K, Nakaya T, Ohta K, Suda T, Rezano A, Kitabatake M, Vaeteewoottacharn K, Okada S, Tone S, et al. 2012. Selective cell death of p53-insufficient cancer cells is induced by knockdown of the mRNA export molecule GANP. *Apoptosis Int J Program Cell Death* **17**: 679–690.

Picard-Jean F, Bougie I, Shuto S, Bisaillon M. 2013. The immunosuppressive agent mizoribine monophosphate is an inhibitor of the human RNA capping enzyme. *PloS One* **8**: e54621.

Pina AS, Hussain A, Roque ACA. 2009. An historical overview of drug discovery. *Methods Mol Biol Clifton NJ* **572**: 3–12.

Plessel G, Fischer U, Lührmann R. 1994. M(3)G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro - evidence that the U1 small nuclear RNA - (guanosine-N2) - methyltransferase is a non-snNRP

cytoplasmic protein that requires a binding site of the Sm core domain. *Mol Cell Biol* **14**: 4160–4172.

Pratt JP, Zeng Q, Ravnic D, Huss H, Rawn J, Mentzer SJ. 2009. Hierarchical clustering of monoclonal antibody reactivity patterns in nonhuman species. *Cytom Part J Int Soc Anal Cytol* **75**: 734–742.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**: 218–29.

Pyne S, Hu X, Wang K, Rossin E, Lin T-I, Maier LM, Baecher-Allan C, McLachlan GJ, Tamayo P, Hafler DA, et al. 2009. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci U S A* **106**: 8519–8524.

Qiu P, Simonds EF, Bendall SC, Gibbs KD Jr, Bruggner RV, Linderman MD, Sachs K, Nolan GP, Plevritis SK. 2011a. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* **29**: 886–891.

Qiu ZR, Shuman S, Schwer B. 2011b. An essential role for trimethylguanosine RNA caps in Saccharomyces cerevisiae meiosis and their requirement for splicing of SAE3 and PCH2 meiotic pre-mRNAs. *Nucleic Acids Res* **39**: 5633–5646.

Raghunathan PL, Guthrie C. 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* **8**: 847–855.

Raijmakers R, Schilders G, Pruijn GJM. 2004. The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm. *Eur J Cell Biol* **83**: 175–183.

Rain JC, Legrain P. 1997. In vivo commitment to splicing in yeast involves the nucleotide upstream from the branch site conserved sequence and the Mud2 protein. *EMBO J* **16**: 1759–1771.

Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, Fedorov A. 2011. Critical association of ncRNA with introns. *Nucleic Acids Res* **39**: 2357–2366.

Reddy R, Busch H. 1988. Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. In *Structure and Function of Major and Minor Small Nuclear*

Ribonucleoprotein Particles (ed. P.D.M.L. Birnstiel), pp. 1–37, Springer Berlin Heidelberg

Rix U, Superti-Furga G. 2009. Target profiling of small molecules by chemical proteomics. *Nat Chem Biol* **5**: 616–624.

Rodriguez-Navarro S, Fischer T, Luo MJ, Antunez O, Brettschneider S, Lechner J, Perez-Ortin JE, Reed R, Hurt E. 2004. Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell* **116**: 75–86.

Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, Fangazio M, Vaisitti T, Monti S, Chiaretti S, et al. 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood* **118**: 6904–6908.

Rutz B, Séraphin B. 1999. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* **5**: 819–831.

Sakharkar MK, Chow VTK, Kangueane P. 2004. Distributions of exon and introns in the human genome. *Silico Biol* **4**: 0032.

Saldanha AJ. 2004. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**: 3246–3248.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Labroatory Press, Plainview, N.Y.

Sanchez A, Golding I. 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**: 1188–1193.

Sartipy P, Björquist P, Strehl R, Hyllner J. 2007. The application of human embryonic stem cell technologies to drug discovery. *Drug Discov Today* **12**: 688–699.

Sayani S, Chanfreau GF. 2012. Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in Saccharomyces cerevisiae. *RNA* **18**: 1563–1572.

Sayani S, Janis M, Lee CY, Toesca I, Chanfreau GF. 2008. Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. *Mol Cell* **31**: 360–370.

Schenone M, Dančík V, Wagner BK, Clemons PA. 2013. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* **9**: 232–240.

Schmidt K, Butler JS. 2013. Nuclear RNA Surveillance: Role of TRAMP in Controlling Exosome Specificity. *Wiley Interdiscip Rev RNA* **4**: 217–231.

Schneider BL, Steiner B, Seufert W, Futcher AB. 1996. pMPY-ZAP: a reusable polymerase chain reaction-directed gene disruption cassette for Saccharomyces cerevisiae. *Yeast Chichester Engl* **12**: 129–134.

Schwer B, Gross CH. 1998. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J* **17**: 2086–2094.

Schwer B, Guthrie C. 1991. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**: 494–9.

Schwer B, Shuman S. 1996. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA* **2**: 574–583.

Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ. 1999. Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A* **96**: 13638–13643.

Segref A, Sharma K, Doye V, Hellwig A, Huber J, Lührmann R, Hurt E. 1997. Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. *EMBO J* **16**: 3256–3271.

Séraphin B, Kretzner L, Rosbash M. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5’ cleavage site. *EMBO J* **7**: 2533–2538.

Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**: 1567–72.

Sharma A, Jacob A, Tandon M, Kumar D. 2010. Orphan drug: Development trends and strategies. *J Pharm Bioallied Sci* **2**: 290–299.

Sheff MA, Thorn KS. 2004. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast Chichester Engl* **21**: 661–670.

Sherman F. 1991. Getting Started with Yeast. *Methods Enzymol* **194**: 3–21.

Shilo A, Ben Hur V, Denichenko P, Stein I, Pikarsky E, Rauch J, Kolch W, Zender L, Karni R. 2014. Splicing factor hnRNP A2 activates the Ras-MAPK-ERK pathway by controlling A-Raf splicing in hepatocellular carcinoma development. *RNA* **20**: 505–515.

Siddiqui N, Borden KLB. 2011. mRNA export and cancer. *Wiley Interdiscip Rev RNA* n/a–n/a.

Sieber SA, Marahiel MA. 2003. Learning from Nature's Drug Factories: Nonribosomal Synthesis of Macrocyclic Peptides. *J Bacteriol* **185**: 7036–7043.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**: 19–27.

Simoes-Barbosa A, Meloni D, Wohlschlegel JA, Konarska MM, Johnson PJ. 2008. Spliceosomal snRNAs in the unicellular eukaryote Trichomonas vaginalis are structurally conserved but lack a 5'-cap structure. *RNA* **14**: 1617–1631.

Simon JA, Bedalov A. 2004. Yeast as a model system for anticancer drug discovery. *Nat Rev Cancer* **4**: 481–487.

Sonenberg N. 1993. Remarks on the mechanism of ribosome binding to eukaryotic mRNAs. *Gene Expr* **3**: 317–323.

Sonenberg N, Hinnebusch AG. 2009. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**: 731–745.

Sorenson MR, Stevens SW. 2014. Rapid identification of mRNA processing defects with a novel single-cell yeast reporter. *RNA* **20**: 732–745.

Spingola M, Grate L, Haussler D, Ares M. 1999. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA* **5**: 221–234.

Srebrow A, Kornblihtt AR. 2006. The connection between splicing and cancer. *J Cell Sci* **119**: 2635–2641.

St Onge R, Schlecht U, Scharfe C, Evangelista M. 2012. Forward chemical genetics in yeast for discovery of chemical probes targeting metabolism. *Mol Basel Switz* **17**: 13098–13115.

Staley JP, Guthrie C. 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* **3**: 55–64.

Stewart M. 2010. Nuclear export of mRNA. *Trends Biochem Sci* **35**: 609–17.

Stockley J, Villasevil MEM, Nixon C, Ahmad I, Leung HY, Rajan P. 2014. The RNA-binding protein hnRNPA2 regulates β-catenin protein expression and is overexpressed in prostate cancer. *RNA Biol* **11**.

Strasser K, Hurt E. 2000. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J* **19**: 410–420.

Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Andres AK, Struhl K, Reed R, et al. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**: 304–308.

Stutz F, Bachi A, Doerks T, Braun IC, Séraphin B, Wilm M, Bork P, Izaurralde E. 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**: 638–650.

Stutz F, Liao XC, Rosbash M. 1993. U1 small nuclear ribonucleoprotein particle-protein interactions are revealed in Saccharomyces cerevisiae by in vivo competition assays. *Mol Cell Biol* **13**: 2126–2133.

Sundberg SA. 2000. High-throughput and ultra-high-throughput screening: solution- and cell-based approaches. *Curr Opin Biotechnol* **11**: 47–53.

Suzuki Y, St Onge RP, Mani R, King OD, Heilbut A, Labunskyy VM, Chen W, Pham L, Zhang LV, Tong AHY, et al. 2011. Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. *Nat Methods* **8**: 159–164.

Swida U, Thuroff E, Kaufer NF. 1986. Intron mutations that affect the splicing efficiency of the CYH2 gene of Saccharomyces cerevisiae. *Mol Gen Genet* **203**: 300–304.

Swida U, Thuroff E, Steinert E, Kaufer NF. 1988. A non-conserved sequence in the 5' region of the CYH2 Intron from Saccharomyces cerevisiae controls splicing efficiency of the pre-mRNA. *Yeast* **4**: 209–217.

Szkukalek A, Myslinski E, Mougin A, Lührmann R, Branlant C. 1995. Phylogenetic conservation of modified nucleotides in the terminal loop-1 of the spliceosomal U5 snRNA. *Biochimie* **77**: 16–21.

Tarun SZ, Sachs AB. 1995. A common function for mRNA 5' and 3' ends in translation initiation in yeast. *Genes Dev* **9**: 2997–3007.

Tavassoli A, Benkovic SJ. 2007. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nat Protoc* **2**: 1126–1133.

Teem JL, Rosbash M. 1983. Expression of a beta-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the rna2 mutation of yeast. *Proc Natl Acad Sci U S A* **80**: 4403–4407.

Tentori L, Leonetti C, Scarsella M, D'Amati G, Vergati M, Portarena I, Xu W, Kalish V, Zupi G, Zhang J, et al. 2003. Systemic administration of GPI 15427, a novel poly(ADP-ribose) polymerase-1 inhibitor, increases the antitumor activity of temozolomide against intracranial melanoma, glioma, lymphoma. *Clin Cancer Res Off J Am Assoc Cancer Res* **9**: 5370–5379.

Teplova M, Wohlbold L, Khin NW, Izaurralde E, Patel DJ. 2011. Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. *Nat Struct Mol Biol* **18**: 990–998.

Thongyoo P, Roqué-Rosell N, Leatherbarrow RJ, Tate EW. 2008. Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. *Org Biomol Chem* **6**: 1462–1470.

Thorstholm L, Craik DJ. 2012. Discovery and applications of naturally occurring cyclic peptides. *Drug Discov Today Technol* **9**: e13–e21.

Tian B, Graber JH. 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* **3**: 385–396.

Titz B, Thomas S, Rajagopala SV, Chiba T, Ito T, Uetz P. 2006. Transcriptional activators in yeast. *Nucleic Acids Res* **34**: 955–967.

Tochigi Y, Sato N, Sahara T, Wu C, Saito S, Irie T, Fujibuchi W, Goda T, Yamaji R, Ogawa M, et al. 2010. Sensitive and convenient yeast reporter assay for high-throughput analysis by using a secretory luciferase from Cypridina noctiluca. *Anal Chem* **82**: 5768–5776.

Tohoyama H, Shiraishi E, Amano S, Inouhe M, Joho M, Murayama T. 1996. Amplification of a gene for metallothionein by tandem repeat in a strain of cadmium-resistant yeast cells. *FEMS Microbiol Lett* **136**: 269–273.

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**: 2364–2368.

Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, et al. 2004. Global Mapping of the Yeast Genetic Interaction Network. *Science* **303**: 808–813.

Torres NP, Lee AY, Giaever G, Nislow C, Brown GW. 2013. A high-throughput yeast assay identifies synergistic drug combinations. *Assay Drug Dev Technol* **11**: 299–307.

Tu D, Li W, Ye Y, Brunger AT. 2007. Structure and function of the yeast U-box-containing ubiquitin ligase Ufd2p. *Proc Natl Acad Sci USA* **104**: 15599–606.

Tucker M, Staples RR, Valencia-Sanchez MA, Muhlrad D, Parker R. 2002. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in Saccharomyces cerevisiae. *EMBO J* **21**: 1427–1436.

Tugendreich S, Perkins E, Couto J, Barthmaier P, Sun D, Tang S, Tulac S, Nguyen A, Yeh E, Mays A, et al. 2001. A streamlined process to phenotypically profile heterologous cDNAs in parallel using yeast cell-based assays. *Genome Res* **11**: 1899–912.

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al. 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**: 623–627.

Vasudevan S, Peltz SW, Wilusz CJ. 2002. Non-stop decay--a new mRNA surveillance pathway. *BioEssays News Rev Mol Cell Dev Biol* **24**: 785–788.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. *Science* **291**: 1304–+.

Du Vigneaud V, Ressler C, Trippett S. 1953. The sequence of amino acids in oxytocin, with a proposal for the structure of oxytocin. *J Biol Chem* **205**: 949–957.

Vijayraghavan U, Company M, Abelson J. 1989. Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. *Genes Dev* **3**: 1206–16.

Volckaert G, Voet M, Van der Schueren J, Robben J, Vanstreels E, Vander Stappen J. 2003. Disruption of 12 ORFs located on chromosomes IV, VII and XIV of Saccharomyces cerevisiae reveals two essential genes. *Yeast Chichester Engl* **20**: 79–88.

Wacker SA, Houghtaling BR, Elemento O, Kapoor TM. 2012. Using transcriptome sequencing to identify mechanisms of drug action and resistance. *Nat Chem Biol* **8**: 235–237.

Waksman SA, Woodruff HB. 1940. Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomyces. *Exp Biol Med* **45**: 609–614.

Wang G-S, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**: 749–761.

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, et al. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* **365**: 2497–2506.

Wang W, Czaplinski K, Rao Y, Peltz SW. 2001. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J* **20**: 880–890.

Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**: 437–440.

Wells SE, Hillner PE, Vale RD, Sachs AB. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* **2**: 135–140.

Werven FJ van, Timmers HTM. 2006. The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Res* **34**: e33–e33.

Wickramasinghe VO, Stewart M, Laskey RA. 2010. GANP enhances the efficiency of mRNA nuclear export in mammalian cells. *Nucl Austin Tex* **1**: 393–396.

Will CL, Lührmann R. 2011. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3**: a003707.

Wilmes GM, Bergkessel M, Bandyopadhyay S, Shales M, Braberg H, Cagney G, Collins SR, Whitworth GB, Kress TL, Weissman JS, et al. 2008. A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol Cell* **32**: 735–746.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**: 901–906.

Wise JA. 1991. Preparation and analysis of low-molecular weight RNAs and small ribonucleoproteins. *Meth Enzym* **194**: 405–415.

Wolffe A. 1998. *Chromatin: Structure and Function*. Academic Press.

Yaffe MP, Schatz G. 1984. Two nuclear mutations that block mitochondrial import in yeast. *Proc Natl Acad Sci USA* **81**: 4819–4823.

Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, et al. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**: 64–69.

Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al. 2008. High-quality binary protein interaction map of the yeast interactome network. *Science* **322**: 104–110.

Yu Y-T, Shu M-D, Steitz JA. 1998. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J* **17**: 5783–5795.

Zenklusen D, Stutz F. 2001. Nuclear export of mRNA. *FEBS Lett* **498**: 150–156.

Zenklusen D, Vinciguerra P, Strahm Y, Stutz F. 2001. The yeast hnRNP-like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol Cell Biol* **21**: 4219–4232.

Zhu Y, Qi C, Cao WQ, Yeldandi AV, Rao MS, Reddy JK. 2001. Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *Proc Natl Acad Sci U S A* **98**: 10380–10385.

Vita

Matthew Richard Sorenson was born in 1984 in Carol Stream, Illinois. After earning his B.S. in Biology from North Central College he moved to Austin, TX to begin his graduate work at The University of Texas at Austin.

mrsorenson@utexas.edu

This dissertation was typed by Matthew Richard Sorenson.